Evolutionary Thinking Across Disciplines

Problems and Perspectives in Generalized Darwinism

Agathe du Crest • Martina Valković
André Ariew • Hugh Desmond
Philippe Huneman • Thomas Reydon

Editors

Synthese Library,

Springer

Table of Contents

Chapter 1

Generalizing Darwinism as a Topic for Multidisciplinary Debate

Agathe du Crest[i], Martina Valković[ii], André Ariew[iii], Hugh Desmond[iv], Philippe Huneman[v] & Thomas A.C. Reydon[vi]

---

[i] Institut d'Histoire et de Philosophie des Sciences et des Techniques, University of Paris I Panthéon-Sorbonne, Paris.

[ii] Institute of Philosophy, Leibniz University Hannover; Radboud University, Nijmegen.

[iii] Department of Philosophy, University of Missouri in Columbia, Missouri.

[iv] Institut d'Histoire et de Philosophie des Sciences et des Techniques (IHPST), University of Paris I Panthéon-Sorbonne, Paris, and the Department of Philosophy, University of Antwerp.

[v] CNRS and Institut d'Histoire et de Philosophie des Sciences et des Techniques, University of Paris I Panthéon-Sorbonne, Paris.

[vi] Institute of Philosophy and the Centre for Ethics and Law in the Life Sciences (CELLS), Leibniz University Hannover.

1.1. The Problem: Generalizing Darwinism

The ideas that were published in Darwin’s *On the Origin of Species* (as well as in the later *The Descent of Man)* continue to have a major impact on our understanding of the world in which we live and the place that humans occupy in it. They constitute the core of the contemporary life sciences, and elicit enduring fascination as a potentially unifying basis for various branches of biology and the biomedical sciences. They can be used to understand the biological ground of cognition and psychopathology in psychology, psychiatry, or neuroscience. Perhaps the best known expression of this fact is Dobzhansky’s famous dictum that “nothing in biology makes sense except in the light of evolution” (Dobzhansky, 1964: 449; 1973: 125), and given that all human behavior supervenes on some biological basis, evolutionary thinking has a vast scope even just in this regard.

However, there has long been the conviction that the significance of evolutionary thinking goes beyond biology. Just consider the motto “survival of the fittest”, Spencer’s pithy redescription of the principle natural selection: the phrase does not specify what entities are at play. There is some such struggle at play not just between organisms, but also cultures, norms, or even ideas. In fact, Thomas Huxley, sometimes better known as “Darwin’s bulldog”, had intuited by 1880 that “the struggle for existence holds as much in the intellectual as in the physical world” (Huxley 1880, 15–16). Here Huxley anticipates a how Darwin’s theory of the “transmutation” of populations of organisms could also be applied to the transmutation of populations of technical artifacts, social institutions, moral norms, or economic entities.

It is difficult to say when such generalizations of evolutionary thinking moved beyond intuitions and were fleshed out in a systematic way. Arguably, after the publication of the *Origin*, it was Darwin himself who realized that, because of its generality, the theory can be extended to explain phenomena beyond the non-human living world. The *Descent of Man* (1871) contained the first generalizations of the ideas in the *Origin* to the human mind and human culture. However, subsequent generations of naturalist thinkers sought to apply Darwin’s theory more broadly and systematically: to the human mind, to human behavior, to human diversity and differences between groups, and to society (Richards, 1987). Prominent early examples include authors such as Herbert Spencer with respect to the foundations of social science, and applications of Darwinian ideas to societal issues (Spencer 1876), William James whose functionalist approach to psychology was inspired by Darwin’s (see Green 2009), Torstein Veblen for economic (Veblen 1898) and John Dewey with respect to pragmatic philosophy and education (Dewey 1910).

Despite this initial growth of evolutionary approaches, it is only in recent decades that there has been an acceleration in the interest in using evolutionary concepts and models to describe and explain non-biological phenomena (for a recent overview, see Heams et al., 2015). The following non-exhaustive list gives a a sense of how broad the variety of evolutionary subfields is:

- evolutionary economics (Nelson & Winter, 1982; Witt, 2003; Hodgson, 2019; Witt & Chai, 2019),
- evolutionary anthropology and cultural evolutionary theory (Boyd & Richerson, 2005; Mesoudi, 2011; Mesoudi et al., 2006),
- evolutionary sociology (Dietz et al., 1990; Hopcroft, 2016; Turner & Machalek, 2018),
- evolutionary psychology (Barkow et al., 1992; Buss, 1995; 2008),
- evolutionary literary studies (Carroll, 2004; Gottschall & Wilson, 2005),
- evolutionary archaeology (Maschner, 1996; Barton & Clark, 1997),
- evolutionary history (Stuart-Fox, 2002; Russell, 2011),
- evolutionary medicine (Nesse & Williams, 1995; Nesse & Stearns, 2008; Stearns, 2012; Rühli & Henneberg, 2013),
- evolutionary computation (Mitchell & Taylor, 1999; Eiben & Smith, 2015),
- evolutionary electronics (Zebulum et al., 2002; Haddow & Tyrrell, 2011),
- quantum Darwinism (Blume-Kohout & Zurek 2006; Zurek 2018)
- evolutionary epistemology (Popper, 1972; Campbell, 1974; Gontier & Bradie, 2021),
- evolutionary ethics (Ruse, 1986; Joyce, 2006),
- evolutionary aesthetics (Voland & Grammer, 2003; Kozbelt, 2017),
- evolution of science and technology (Hull, 1980; 1988; Ziman, 2000; Brey, 2008; Mesoudi et al., 2013; Scerri, 2016).

Some evolutionary approaches are primarily manifested as theoretical possibilities in journal publications. For instance, Quantum Darwinism is based on the idea that the collapse of the wave function is interpreted as a type of "natural selection" between the "fittest" quantum states. While this work is still ongoing, it seems fair to say that it has not burgeoned into a proper subfield of quantum physics. This contrasts with other evolutionary approaches, with evolutionary psychology and evolutionary anthropology as prime examples. These approaches are much older, arguably originating with Darwin's *Descent* (1871), and have by today grown into subfields with all the corresponding sociological hallmarks: scientific journals, scientific associations, and even departments dedicated to the subfield.

Another distinguishing factor between these evolutionary approaches is that they do not all use evolutionary thinking for the same type of theoretical purpose. When evolutionary thinking is applied in philosophy – ethics or epistemology – the reason is that it can offer an analysis of moral norms or the concept of knowledge in a way that is wholly naturalistic, i.e., in a way that makes no reference to reasons or rationality. Such evolutionary approaches in philosophy typically elicit strong resistance from more traditional ethicists or epistemologists, who view evolutionary approaches as succumbing to the naturalistic fallacy where reasons and causes are confused.

By contrast, for the social sciences, and perhaps some of the humanities such as history, the attraction of evolutionary thinking is the hope that it can offer a solid theoretical and possibly even unifying foundation for a variety of branches of inquiry (Mesoudi et al., 2006). Historiography has traditionally been conceived of more as an art than as a science, with the core business of historians to weave narratives that help make sense of events after the fact. The idea that historians would attempt to construct predictive theories was once deemed to be so misguided as to only produce "intellectual monsters" (Danto [1985]2007: 15). Today, by contrast, scientific approaches to history such as cliodynamics are gaining traction, and with it, the attractiveness of evolutionary approaches to history has grown, despite reluctance of traditional historians (e.g. Turchin 2008).

This overview could give the impression that the history of the reception of evolutionary thinking has been one of a steadily increasing popularity. This has emphatically not been the case: evolutionary thinking outside biology remains contested, and has been so since its very inception. In the academic sphere, skepticism may take a more implicit or silent form, namely as a lack of enthusiasm or interest in evolutionary approaches. For instance, in economics, evolutionary approaches were first formulated more than a century ago with Veblen (1898), were picked up again by Nelson and Winter (1982) and more recently have found a strong advocate in the work of Hodgson and Knudsen (2010). Nonetheless, the economics that is considered "mainstream" receives epithets such as "Keynesian" or "Friedmanite" – not

“Darwinian”. Evolutionary economics is still categorized among the “alternative” or “heterodox” approaches in economics, together with “institutional economics” or “ecological economics“.

Economics is an interesting test case to reflect on why evolutionary approaches should have stronger uptake in some domains of inquiry rather than others. Evolutionary economics is one of the oldest generalizations of evolutionary thinking to other domains, and so contradicts the thesis that only younger evolutionary approaches are the less well-established ones. Moreover, it is not that evolutionary thinking has had no influence on economics. Its indirect influence is especially noticeable in behavioral economics (following Tversky and Kahneman 1974) that relies on research evolved cognitive biases. And of course, once one delves deeper, economics and evolutionary theory become even more intertwined. Darwin himself was heavily indebted to the earlier work of economists such as Thomas Malthus and Adam Smith. Later developments in evolutionary thinking, such as research into the evolution of animal behavior (following Maynard Smith and Price 1973), were strongly inspired by game theory. And the structural similarity or at least analogical similarity between core economical and evolutionary concepts seems impossible to miss: competition; the market (vs. environment); utility (vs. fitness); learning (vs. adaptation). Nonetheless, despite this long history of cross-fertilization, it seems fair to say that the standard type of explanation present in standard introductory textbooks on economics (see e.g. McConnell & Brue 2008) continues to be more structurally similar to Galilean mathematical explanations with a small number of precisely definable variables (supply, demand, interest rate, money supply, etc.), rather than to Darwinian explanations involving complex, variable populations interacting with the environment and changing over time. In comparison to psychologists or anthropologists, fewer economists seem to have judged evolutionary thinking to be explicitly indispensable for their explanatory goals. Why precisely this is so, and in general, why evolutionary approaches have had stronger uptake in some domains of inquiry rather than others, remains an open question.

If we move on from the sphere of academic and scientific enquiry, and consider the reception of evolution and Darwinism in the public sphere, this is where we find the loudest critical voices. The oldest and perhaps still today the most widely known example of skepticism towards evolutionary theory in public discourse is the debate between religiously motivated creationist views and proponents of evolutionary thinking. Creationist views can range from mere metaphysical possibilities – such as a hands-off Deism where divine intervention is limited to setting the universe in motion – to full-blown empirical theories such as Young Earth Creationism which purports to explain everything that Darwin's theory explained (biogeography, embryology, distribution of fossils, etc.) but only in reference to a conscious and intentional "Intelligent Designer" (as opposed to natural selection).

This debate was the most prominent in Darwin's day. The idea that humans could have descended from an ape-like ancestor was almost immediately met with a mixture of derision and alarm. However, the origin of creationism also illustrates a fascinating if disturbing way in which evolutionary thinking was generalized in the political sphere. For the strong literalist reading of Genesis is a phenomenon of the late-19th century, and of U.S. evangelical religion in particular. Why this was so is a complex historical phenomenon we do not need to delve in, but one factor that historians point to is that creationists' blanket rejection of Darwin had to do with the perceived connection between Darwinian thinking and eugenics of that era. For the proponents of eugenics – which included many leading evolutionary scientists such as Galton or Fisher – presented their policy proposals as based on evolutionary science. Religiously motivated resistance to eugenics (and Kevles reports that most of the organized resistance to eugenics in the U.S. was mounted by the Catholic Church: Kevles 1985, pp. 118-119) was dismissed as short-sighted sentimentalism and misplaced altruism.

There of course would be much more to be said about the complex connection between evolutionary theory, creationism, and eugenics. But it illustrates how scepticism towards evolutionary thinking was inspired by political and ethical considerations even in the late 19th century. And the perception that evolutionary thinking has political implications has remained strong to this day. Indeed, the generality of Darwin's theory has been said to make it a "dangerous idea" (Dennett, 1995), not just because it has the potential to upset traditional theories and revolutionize fields of science but also because it has a strong potential to be misused in the societal arena.

Nowhere has this been more on display than in debates about the genetic basis of evolutionary change. The old eugenicists were strong genetic determinists, believing that genetic material determined a person's cognitive abilities and temperament, but this view was soon relegated to the status of pseudoscience – as the field of genetics developed after the rediscovery of Mendel's work by Hugo De Vries in 1900, it was soon realized there is a strong environmental component to how genotypes are expressed in phenotypes. However, the idea that evolutionary theory could be used for betterment of the human species remained. One of the foundational texts of the Modern Synthesis, Fisher's Genetical Theory of Natural Selection (1930), not only succeeds in combining Mendelism with the theory of natural selection (with the help of statistics), but also goes on, in the second half of the book, to apply this new theoretical understanding to further eugenic ends.

The current consensus follows Lewontin (1974) cautioning against interpreting measures of heritability as evidence for genetic heritability. Nonetheless, debates about the relative contribution of genes to development of human phenotypes, and especially for certain human properties such as intelligence, remain very sensitive today, especially because the echoes of the excesses of eugenics still resound. In the 1970s and 1980s a debate erupted following the publication of E.O. Wilson's book, *Sociobiology* (Wilson, 1975a; for the debate, see Allen et al., 1975; Wilson, 1975b; 1976; Sociobiology Study Group of Science for the People, 1976; Segerstrale, 1986; Wilson & Wilson, 2007). Wilson, an entomologist by training, refocused attention on the biological and especially genetic basis of human behavior – a focus that elicited criticisms of genetic determinism. Richard Hernnstein and Charles Murray's *The Bell Curve* in the 1980s and its statements about race and intelligence is an infamous example of the political and ethical implications certain types of evolutionary thinking can be perceived to have. Gould's response, *The mismeasure of man* (1982) remains a landmark for the critiques of abuses of genetics in social sciences .More recently, the advent of genome-wide association studies (GWAS), where not single genes but clusters of genes are linked to traits, has arguably given new impetus to more sophisticated forms of genetic determinism (see e.g. Comfort 2018).

Another important example of how skepticism towards generalizations of Darwinism assume political-ethical dimensions is the advent of evolutionary psychology. In its standard form, evolutionary psychology models human cognition in analogy to a Swiss Army knife: consisting of modularized and automatic mental functions, shaped over tens to hundreds of thousands of years by natural selection (Barkow, Cosmides, & Tooby 1992). The field has been criticized on scientific and conceptual grounds over the years, including by philosophers of science (Dupré 2000; 2010; Buller, 2006; Downes 2017; Smith, 2020) questioning whether the evolutionary explanations constructed by some evolutionary psychologists actually constitute good scientific explanations. Other criticisms are ethical in nature, since modelling human cognition in this way – and in particular the modelling of some sex differences in particular – could give rise to forms of scientific sexism. In this way, it is feared that evolutionary explanations of human behavior resting on unwarranted assumptions about human evolution could have profoundly adverse effects on society.

This state of affairs, where the remit evolutionary thinking has expanded well beyond its initial scope, and where evolutionary thinking continues to be perceived as having ethical or political import, gives rise to a number of questions that remain unanswered and that make "generalizing Darwinism" a topic for study and debate for philosophers of science and philosophically minded scientists. First, it raises questions on the nature of scientific explanation and method, about the reduction of one domain to another and, more generally, about the relation between a theory in a given domain and its application to antoher: should one speak of reduction, or metaphor, analogy, isomorphisms, continuity, and so on? Second, it triggers questions about what role values (both epistemic and non-epistemic) play in science. Given the complex history of the reception of evolutionary thinking, as well as the different guises that "Darwinism" has assumed over time, there is also a strong historical component involved with reflecting about generalizations of Darwinism (for more discussion, see Desmond et al. ms).

One of the most fundamental open issues concerns what exactly the scope of application of evolutionary thinking is, and what it can be. The shared epistemological outlook of the evolutionary approaches mentioned above is that Darwinism can assume the role of a multi-purpose explanatory framework, or even a unifying paradigm, that could bring a diversity of fields of investigation both within and outside the life sciences together under a common framework. The shared epistemological outlook raises further questions whether there is also a shared ontology grounding that outlook. In that case, various phenomena, processes and systems under study all instantiate the same basic process. Some have advocated for some nuclear form of Darwinian evolutionary theory that is to be filled in on a case-by-case basis for application to the various phenomena under study (cf. Aldrich et al. 2008; Hodgson & Knudsen 2010). "Paradigm" here need not only refer to paradigms in the Kuhnian sense, but also to other notions such as Lakatosian research programs, Foucauldian *episteme's* or Holtonesque themata: all are candidate notions to capture the status of Darwinism as something more than a theory.

To what extent are such epistemological views and metaphysical assumptions warranted? Are so-called "evolutionary" processes in societies and economies sufficiently similar to biological evolutionary processes to be explained in the same way, or at least in ways that allow them to be subsumed under an overarching evolutionary account? Many evolutionary approaches outside the life sciences rest on analogies between biological evolutionary processes and processes in society, such that a question is how strong the similarities between the various processes must be to enable the formulation of genuinely evolutionary explanations in the social and cultural domain (Mayntz, 1992).

An important aspect of this question is what a complete evolutionary explanation of a biological phenomenon exactly should encompass: what is the structure of evolutionary explanations and what components should a good evolutionary explanation have (Reydon, 2022)? And if evolutionary explanations *can* be formulated in different areas of investigation, what follows for the unity and structure of the sciences – in particular, to what extent would evolutionary thinking be able to serve as a unifying paradigm for the life and social sciences, as some authors suggest (Mesoudi et al., 2006) and others have cautiously doubted (Reydon, 2021)?

A second cluster of problems arises from applying evolutionary thinking to human and societal phenomena in particular. There is a difference, in this regard, between applying evolutionary thinking to – for instance – differences between male and female mate choice and applying the same thinking to the collapse of wave functions. Understanding why precisely some generalizations but not others are perceived to be politically charged is not an easy task. Some theories such as the second law of thermodynamics, including its information-theoretic formulation, seems to be at least as generalizable. However, it seems reasonable to presume that any application of concepts such as entropy or mutual information to human behavior would not be perceived to have the same charged consequences that applications of natural selection are perceived to have.

A final cluster of problems that could threaten to bring any enquiry back to square one, is whether, even within the relatively narrow context of biology, it is unambiguously clear what precisely "evolutionary thinking" entails (see also Desmond et al. ms). The ideas introduced by Darwin underwent at least one major transformation in the 1920s and 1930s through the rise of the Modern Synthesis, which sought to unify natural selection with Mendelism by means of statistics (and many of the founders of the Modern Synthesis, such as Fisher, Wright, or Pearson, are also key figures in modern statistics). This yielded a precise mathematical approach to evolution, but which has been criticized for foregrounding genes and alleles and moving processes at the level of the organism, such as development, metabolism, or behavior, to the background (Bateson 2005; Walsh 2015). Recently the "Extended Evolutionary

Synthesis" has been advancing an alternative view where the organism, through plasticity or niche construction, plays a more active role. Just how distinct the Extended Synthesis is from the Modern Synthesis, and to what extent it generates new predictions, remains a topic of active discussion. However, it does pose the question what precisely is being generalized when evolutionary approaches are advanced in new fields.

1.2. Aims, Structure and Content of the Book

The present volume aims to advance philosophical and theoretical work on these questions by providing a survey of the current state of the art in research on relevant topics in the life sciences, the philosophy of science, and the various areas of evolutionary research outside the life sciences. The volume aims to achieve more clarity on the epistemic potential of applying evolutionary thinking outside biology. By bringing together chapters by evolutionary biologists, systematic biologists, philosophers of biology, philosophers of social science, complex systems modelers, psychologists, anthropologists, economists, linguists, historians, and educators, the volume examines evolutionary thinking within and outside the life sciences from a multidisciplinary perspective.[1]

While the chapters written by biologists and philosophers of science address theoretical aspects of the guiding questions and aims of the volume, the chapters written by researchers from the other areas approach them from the perspective of applying evolutionary thinking to non-biological phenomena. Taken together, the chapters in this volume do not only show how evolutionary thinking can be fruitfully applied in various areas of investigation, but also

---

[1] The volume originates in the expert workshop "Evolutionary Thinking Across Disciplines. Problems and Perspectives in Generalized Darwinism", which was organized by the volume editors at the *Institut des Systèmes Complexes* in Paris in October 2021. This expert workshop was the first event that was organized in the context of the project "The Explanatory Scope of Generalized Darwinism: Towards Criteria for Evolutionary Explanations Outside Biology" (GenDar), a research project located at the *Institut d'Histoire et de Philosophie des Sciences et des Techniques*, *CNRS / Université Paris I Panthéon-Sorbonne*, and the *Institut für Philosophie*, *Leibniz Universität Hannover*, and jointly funded by the *Agence Nationale de la Récherche* (ANR) and the *Deutsche Forschungsgemeinschaft* (DFG). The GenDar project closely collaborates with the Evolution and Social Science group at the University of Missouri, and the expert workshop in Paris was a joint event.

highlight numerous open problems, unanswered questions, and issues on which more clarity is needed. As such, the volume can serve as a starting point for future research on the application of evolutionary thinking across disciplines.

There are many ways to cluster the various contributions to reveal emergent themes. The following may be useful as a guide for the reader. We highlight four thematic clusters in the volume (that are reflected in the volume's table of contents).

Part I: How can disciplines benefit from, or contribute to, evolutionary frameworks?

Psychology: In "Is a Non-Evolutionary Psychology Possible?" (Chapter 2), Daniel Nettle and Thom Scott-Phillips argue that the contentious debates in human psychology about the viability of the Evolutionary Psychology as an alternative to mainstream human psychology is based upon a false evolution/non-evolutionary dichotomy. Most if not all psychology investigates the functional design in organismal structures. And, since the presence of functional design pre-supposes an origin in evolution by natural selection, most if not all of psychology is evolutionary. The upshot is that psychologists can endorse evolutionary approaches to their investigations without necessary commitments to the most contentious parts of canonical EP.

Economics: In "Evolutionary Economics and the Theory of Cultural Evolution" (Chapter 3), Ulrich Witt argues that cultural evolutionists could help explain how innovative adaptations arise out of groups of agents by providing a theory of how individuals are motivated to search and adopt new activities. In exchange, economic theories could aid cultural evolutionists in explaining the prominence of intentionality. Witt proposes that human motivation serves as the mechanism for innovative expansion of resources which, in turn, produces a bias in the selective population that drives the evolution of innovative adaptations in the economy. In this way human motivations and intentions are the forces that shape innovative adaptations.

Humanities and literary studies: In "Repetition without Replication: Notes Towards a Theory of Cultural Adaptation" (Chapter 4), Carsten Strathausen argues that theoretical insights from extended evolutionary studies in the sciences and cultural adaptation studies in the humanities should serve as a basis for a theoretical framework for the study of cultural adaptation. Strathausen argues that previous attempts have been by neo-Darwinian on the scientific side, and a long-standing bias against statistical-quantitative approaches to culture, on the humanities side. Strathausen proposes replacing both traditions with a relationist approach inspired by the Extended Evolutionary Synthesis that analyzes the dynamic network of interrelated products, processes, and receptions by which artistic material is continually refitted into different forms for new audiences.

In "The Epistemological and Ideological Stakes of Literary Darwinism" (Chapter 5), Alexandre Gefen investigates the debates between humanist practitioners of cultural studies and literary Darwinists who advocate naturalizing literary aesthetic practices. Gefen finds that the virulent criticisms of evolutionary theorists hinder a serene examination of the disciplinary proposals put forward, in all their richness and epistemological ambitions.

Linguistics: In "Evolutionary Aspects of Language Change" (Chapter 6), linguist Johann-Mattis List presents important evolutionary aspects of language change which has not been adequately communicated across other disciplines that investigate human evolution. List then exposes current challenges of evolutionary studies in historical linguistics in light of these evolutionary aspects.

Education sciences: In "A Community Science Model for Interdisciplinary Evolution Education and School Improvement" (Chapter 7), education scientists Dustin Eirdosh and Susan Hanisch argue that a generalized conception of evolution provides a potential for engaging students in understanding the cultural evolutionary dynamics of their everyday lives, schools, and broader communities. They describe a model of Community-Based Cultural Evolution for inter-institutional collaboration at the intersection of evolution education and applied school improvement efforts. Their paper provides a conceptual foundation for exploring the claim that engaging students in reflecting on the cognitive, behavioral, and cultural evolutionary processes in their everyday lives provides new opportunities for school improvement and interdisciplinary evolution education initiatives.

In a follow-up chapter, "Teaching for the Interdisciplinary Understanding of Evolutionary Concepts" (Chapter 8), Eirdosh and Hanisch explore how understanding the structures of knowledge, or the organization of facts and generalizations in science, cognition, and education, may help illuminate the educational potential and evidence-informed pedagogical practices appropriate for teaching about the interdisciplinary application of evolutionary concepts.

Part II: Generalizations of evolutionary theory: common principles or explanatory structures?

Mathematical models: In “From Games to Graphs: Evolving Networks in Cultural Evolution” (Chapter 9), Karim Baraghith proposes a multi-level representation of cultural evolution that takes into account the various entities that evolve in cultures, from interactions between agents to the whole social networks that themselves evolve. Baraghith’s representation is in the form of a mathematical model that draws upon game theory for representing the micro-level interactions and graph theory for the cultural macrolevel.

Ontological commitments: Gerhard Schurz in “Metaphysics of Evolution: Ontology and Justification of Generalized Evolution Theory” (Chapter 10) argues that the key for a successful abstraction and hence subsumption of cultural evolution into the domain of evolutionary theory is the common ontological commitments between biological and cultural evolution in terms of the entities and structures postulated: self-reproducing systems with variation and in which populations are located in environments with limited resources. The main difference between evolution and cultural evolution is that in the former genes are “constructors” of evolutionary systems while in the latter memes are “modificators”.

Multi-level analysis: In “Human Social Evolution via Four Coevolutionary Levels” (Chepter 11), Ted Koditschek proposes a novel framework for a new paradigm of investigating social evolution in scientifically defensible terms. The framework involves four analytically distinct but empirically nested levels (and logics) of evolution: a biological level that adopts the logic of natural selection, a cultural level that advocates a non-reductive logic of cultural selection), a political level, grounded in a ‘logic of domination’, and an economic level, driven by a ‘logic of capitalist competition’ to track human evolutionary history up to the present day. Koditschek follows his account with a general assessment of the intellectual benefits that such retrodictive accounts of longue durée evolutionary history might bring to the social sciences.

Part III: Why should we be skeptical of generalizations of Darwinism?

Limitations for a generalized Darwinism for physics and chemistry: In “Is Natural Selection Physical?” (Chapter 17), Sylvain Charlat, Thomas Heams, and Olivier Rivoire ask whether evolution by natural selection could be applied to understand physico-chemical systems including the origins of life? They argue that in its common organismal formulations, it cannot because it fails to recognize that biological evolution is a process that recursively modifies its own modes of operation. They provide attempts at a resolution.

Epistemic and moral risk: In “The Risks of evolutionary explanation” (Chapter 18), anthropologist H. Clark Barrett considers two kinds of risk in posting evolutionary explanations, the epistemic risk of providing false explanations and the ethical risk involved when those false explanations cause harm. Barrett warns that the allure of evolutionary explanations along with an academic incentive structures makes evolutionists particularly at risk of endorsing false explanations that have the potential for ethical harm.

Limits of Darwinian economics: In “Evolution and Ecology of Organizations and Markets” (Chapter 19), economist Randall E. Westgren examines the prospect of locating the evolution of economic organizations and markets within Generalized Darwinism though the investigation of Joseph Schumpeter’s model of economic evolution form his 1939 book, Business Cycles which features a complex combination of competition and selection processes within a community ecology of economic mutualisms, conventions, institutions, and other inter-firm structures that confound the explanation of the success of the variants. Westgren concludes that the selection and retention features of current accounts of Generalized Darwinism are incompatible with organizational evolution.

Limits of cultural evolutionary theory: Philosopher Simon Lohse in “Pluralism and Epistemic Goals: Why the Social Sciences Will (Probably) Not Be Synthesised by Evolutionary Theory” (Chapter 20) provides a critical assessment of a movement in the literature – lead primarily by Mesoudi and his colleagues – to synthesize the social sciences based on a theory of cultural evolution. Lohse proposes problems with the theory of cultural evolution that Mesoudi and others have proposed, and questions the epistemic value added to social sciences that their evolutionary approach offers.

Methodological limitations: In “Equations at an Exhibition: On the Cultural Price Equation” (Chapter 21), philosopher Tim Lewens considers the degree to which the Price equation serves as a useful analytic tool for the investigation of the evolution of culture. Lewens points out that in cases where a non-distorted distinction between selection and transmission cannot be made, the Price Equation is a misleading analytical tool. Unfortunately, the processes of cultural reproduction make a non-distorted distinction difficult. Hence, the Price Equation can mislead about cultural evolution.

Methodological limitations: In "Unlike Agents: The Role of Correlation in Economics and Biology" (Chapter 22), philosopher Hannah Rubin provides a cautionary tale in adopting ideas about evolution in biology and learning in economics, in particular, the tendency to think of measures of correlation as akin to attitudes of economic agents. The incautious practice leads to use of unreliable heuristics and misunderstandings in biology, as well as to misuse of biological results in economics.

Part IV: How can evolutionary approaches or the target field be amended?

Revisions to the conception of inheritance: In "From the Modern Synthesis to the Inclusive Evolutionary Synthesis: an Einsteinian revolution in evolution" (Chapter 12), Biologist Étienne Danchin argues that recent developments in DNA sequencing and techniques that link DNA variation with phenotypic variation have revealed limitations to the mid-20th century "Modern Synthesis" conception of inheritance. It fails to incorporate the various genetic and non-genetic processes that are part of the inheritance system and hence provides an inadequate view of the full complexity of living organisms. Danchin provides an update which he calls the Inclusive Evolutionary Synthesis. His chapter reflects on historical developments and philosophical reflections on the 21st century science of inheritance as well as personal reflections about the challenges of endorsing IES.

Revisions to the evolutionary theory of development: In "Darwinian/Hennigian Systematics and Evo-Devo: The Missed Rendez-vous" (Chapter 13), Guillaume Lecointre argues that current formulations of "evo devo" are insufficient foundations for the study of morphological complexity of organisms. The field of Evo devo suffers from a false view that genes control body plans and an insufficient regard to the investigation of ontogenetic timing. Lecointre constructs a hierarchical graph of ontogenetic time segments which indicates when organs or other biological structures are present or absent. He argues such "ontophylogenetic" graphs are the real phylogenies that should be at the core of evo devo.

Incorporating a concept of agency: Philosopher Hugh Desmond in his “The Generalized Selective Environment” (Chapter 14), argues that a successful program of generalizing Darwinism to human social activities requires an answer to the question, what constitutes the “selective environment” to which scientific ideas, moral norms, or corporations adapt? A successful answer provides a matter of degree contrast between natural selection and human agency. Attending to the features of the contrast help eliminate conceptual confusions running through the literature.

Incorporating a concept of agency: In “Adding Agency to Tinbergen’s Four Questions” (Chapter 15), philosopher André Ariew and anthropologist Karthik Panchanathan argue that a large part of the lasting appeal to Tinbergen’s four questions was (and still is) the methodological commitment to treating organisms as objects as opposed to purposive agents. While these features are still prized among today’s biological social scientists, it ignores an important feature of many social organisms, that they are not merely objects, they are also purposive agents. Updating Tinbergen’s four questions with agency in mind only makes them more applicable to the biological investigation of animal behavior, but it also strengthens the value and applicability of biology-oriented research programs in the social sciences.

Incorporating human behavioral ecology: In “Cultural Evolution Needs Human Behavioural Ecology and Cultural Transmission Approaches” (Chapter 16), Alberto J. C. Micheletti, Eva Brandl, Hanzhi Zhang, Sarah Peacey, and Ruth Mace employ Tinbergen’s four question framework for the study of behavior and several case studies to distinguish between the questions that human behavioral ecologists answer from those who investigate cultural transmission. They assert that the field of cultural evolution can move forward and achieve greater synthesis by exploring how selective processes acting on biological fitness differ from those acting on cultural fitness – and how the two might interact in the cultural evolution of human behaviours.

References

Aldrich, H.E., Hodgson, G.M., Hull, D.L., Knudsen, T., Mokyr, J. & Vanberg, V.J. (2008): ‘In defence of generalized Darwinism, *Journal of Evolutionary Economics* 18: 577-596.

Allen, E. et al. (1975): ‘Against “sociobiology”’, *The New York Review of Books* 22 (13 November 1975): 43-44.

Barkow, J., Cosmides, L. & Tooby, J. (1992): *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, Oxford: Oxford University Press.

Barton, C.M. & Clark, G.A. (Eds) (1997): *Rediscovering Darwin: Evolutionary Theory and Archeological Explanation (Archeological Papers 7)*, Arlington, VA: American Anthropological Association.

Bateson, Patrick. 2005. "The Return of the Whole Organism." *Journal of Biosciences* 30 (1): 31–39.

Blume-Kohout, Robin, and W. H. Zurek. 2006. "Quantum Darwinism: Entanglement, Branches, and the Emergent Classicality of Redundantly Stored Quantum Information." *Physical Review A* 73 (6): 062310.

Boyd, R. & Richerson, P. (2005): *Not By Genes Alone: How Culture Transformed Human Evolution*, Chicago: University of Chicago Press.

Brey, P. (2008): 'Technological design as an evolutionary process', in:, Vermaas, P.E., Kroes, P., Light, A. & Moore, S.A. (Eds): *Philosophy and Design: From Engineering to Architecture*, Dordrecht: Springer, pp. 61-75.

Buller, D.J. (2006): *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*, Cambridge, MA: MIT Press.

Buss, D.M. (1995): 'Evolutionary psychology: A new paradigm for psychological science', *Psychological Inquiry* 6: 1-49.

Buss, D.M. (2008): *Evolutionary Psychology: The New Science of the Mind*, Boston, MA: Pearson.

Campbell, D.T. (1974): 'Evolutionary epistemology', in: Schilpp, P.A. (Ed.): *The Philosophy of Karl R. Popper*, LaSalle, IL: Open Court, pp. 412-463.

Carroll, J. (2004): *Literary Darwinism: Evolution, Human Nature, and Literature*, New York & London: Routledge.

Comfort, Nathaniel. 2018. "Genetic Determinism Redux." *Nature*, no. 561: 461–63.

Danto, A.C. ([1985] 2007): Narration and Knowledge. New York: Columbia University Press.

Dennett, D.C (1995): *Darwin's Dangerous Idea: Evolution and the Meanings of Life*, Harmondsworth: Penguin.

Dewey, J. (1910): Influence of Darwin on Philosophy and Other Essays. New York: Henry Holt and Company.

Dietz, T., Burns, T.R. & Buttel, F.H. (1990): ‘Evolutionary theory in sociology: An examination of current thinking’, *Sociological Forulitm* 5: 155-171.

Dobzhansky, T. (1964): ‘Biology, molecular and organismic’, *American Zoologist* 4: 443-452.

Dobzhansky, T. (1973): ‘Nothing in biology makes sense except in the light of evolution’, *The American Biology Teacher* 35: 125-129.

Downes, S.M. (2017): ‘Evolutionary psychology’, in: Zalta, E.N. (Ed.): *The Stanford Encyclopedia of Philosophy (Winter 2017 Edition)*, https://plato.stanford.edu/archives/win2017/entries/evolutionary-psychology/.

Dupré, J. (2000): ‘What the theory of evolution can’t tell us’, *Critical Quarterly* 42: 18-34.

Dupré, J. (2010): ‘Causality and human nature in the social sciences’, in: Albert, G. & Sigmund, S. (Eds): *Soziologische Theorie kontrovers*, *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *Sonderheft* 50/2010, pp. 507-525.

Eiben, A.E. & Smith, J. (2015): ‘From evolutionary computation to the evolution of things’, *Nature* 521: 476-482.

Fisher, R. A. (1930) *The Genetical Theory of Natural Selection*. Oxford, UK: Oxford University Press.

Gontier, N. & Bradie, M. (2021): ‘Evolutionary Epistemology: Two research avenues, three schools, and a single and shared agenda’, *Journal for General Philosophy of Science* 52: 197-209.

Gottschall, J. & Wilson, D.S. (Eds) (2005): *The Literary Animal: Evolution and the Nature of Narrative*, Evanston, IL: Northwestern University Press.

Gould, S. J. (1982). *The Mismeasure of Man*. W. W. Norton & Company.

Green, Christopher D. (2009): “Darwinian Theory, Functionalism, and the First American Psychological Revolution.” *American Psychologist* 64 (2): 75–83.

Haddow, P.C. & Tyrrell, A.M. (2011): ‘Challenges of evolvable hardware: past, present and the path to a promising future’, *Genetic Programming and Evolvable Machines* 12: 183-215.

Heams, T., Huneman, P., Lecointre, G. & Silberstein M. (Eds.) (2015): *Handbook of Evolutionary Thinking in the Sciences,* Dordrecht: Springer.

Hodgson, G.M. (2019): *Evolutionary Economics: Its Nature and Future*, Cambridge: Cambridge University Press.

Hodgson, G.M. & Knudsen, T. (2010): *Darwin's Conjecture: The Search for General Principles of Social and Economic Evolution*, Chicago & London: University of Chicago Press.

Hoftstadter, R. (1955): *Social Darwinism in American Thought*, Boston: Beacon Press.

Hopcroft, R. (2016): 'Grand challenges in evolutionary sociology and biosociology', *Frontiers in Sociology* 1: 2.

Hull, D.L. (1980): 'Individuality and selection', *Annual Review of Ecology and Systematics* 11: 311-332.

Hull, D.L. (1988): *Science As a Process: An Evolutionary Account of the Social and Conceptual Development of Science*, Chicago & London: University of Chicago Press.

Joyce, R. (2006): *The Evolution of Morality*, Cambridge, MA: MIT Press.

Kozbelt, A. (2017): 'Contemporary evolutionary aesthetics: The view from the humanities (and humanists)', *Evolutionary Studies in Imaginative Culture* 1: 95-104.

Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

Maschner, H.D.G. (1996): *Darwinian Archaeologies*, New York: Plenum Press.

Maynard Smith, J. & Price, G. R. (1973): "The Logic of Animal Conflict." *Nature* 246 (5427): 15–18.

Mayntz, R. (1992): 'The influence of natural science theories on contemporary social science', in: Dierkes, M. & Biervert, B. (Eds): *European Social Science in Transition: Assessment and Outlook*, Frankfurt am Main: Campus Verlag & Boulder, Co: Westview Press, pp. 27-79.

McConnell, Campbell R., and Stanley L. Brue. 2008. *Economics: Principles, Problems, and Policies*. 17th ed. Boston: McGraw-Hill Irwin.

Mesoudi, A. (2011): *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*, Chicago: University of Chicago Press.

Mesoudi, A., Whiten, A. & Laland, K.N. (2006): 'Towards a unified science of cultural evolution', *Behavioral and Brain Sciences* 29: 329-383.

Mesoudi, A., Laland, K.N., Boyd, R. et al. (2013): 'The cultural evolution of technology and science', in: Richerson, P.J. & Christiansen, M.H. (Eds): *Cultural Evolution: Society, Technology, Language, and Religion*, Cambridge, MA: MIT Press, pp. 193-216.

Mitchell, M. & Taylor, C.E. (1999): 'Evolutionary computation: An overview', *Annual Review of Ecology and Systematics* 30: 593-616.

Nelson, R.R. & Winter, S.G. (1982): *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.

Nesse, R.M. & Stearns, S.C. (2008): 'The great opportunity: Evolutionary applications to medicine and public health', *Evolutionary Applications* 1: 28-48.

Nesse, R.M. & Williams, G.C. (1995): *Evolution and Healing: The New science of Darwinian Medicine*, London: Weidenfeld & Nicolson.

Popper, K.R. (1972): *Objective Knowledge: An Evolutionary Approach*, Oxford: Clarendon Press.

Reydon, T.A.C. (2021): 'Generalized Darwinism as modest unification', *American Philosophical Quarterly* 58: 79-93.

Reydon, T.A.C. (2022): 'The proper role of history in evolutionary explanations', *Noûs*, in press (online advance article available).

Richards, R.J. (1987): *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*, Chicago: University of Chicago Press.

Rühli, F.J. & Henneberg, M. (2013): 'New perspectives on evolutionary medicine: The relevance of microevolution for human health and disease', *BMC Medicine* 11: 115.

Ruse, M. (1986): 'Evolutionary ethics; A phoenix arisen', *Zygon* 21: 95-112.

Russell, E. (2011): *Evolutionary History: Uniting History And Biology To Understand Life On Earth*, Cambridge: Cambridge University Press.

Scerri, E. (2016): *A Tale of Seven Scientists and a New Philosophy of Science*, New York: Oxford University Press.

Segerstrale, U. (1986): ‘Colleagues in conflict: An ‘in vivo’ analysis of the sociobiology controversy’ *Biology & Philosophy* 1: 53-87.

Smith, S.E. (2020): ‘Is evolutionary psychology possible?’, *Biological Theory* 15: 39-49.

Stearns, S.C. (2012): ‘Evolutionary medicine: Its scope, interest and potential’, *Proceedings of the Royal Society B* 279: 4305-4321.

Stuart-Fox, M. (2002): ‘Evolutionary theory of history’, in: Pomper, P. & Shaw, D.G. (Eds): *The Return of Science: Evolution, History, and Theory*, Lanham, MD: Rowman & Littlefield, pp. 123-144.

Barkow, Jerome H., Leda Cosmides, and John Tooby. 1992. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press.

Spencer, H. 1876. The Principles of Sociology. London: Williams & Norgate

Turchin, Peter. 2008. “Arise ‘Cliodynamics.’” *Nature* 454 (7200): 34–35.

Tversky, Amos, and Daniel Kahneman. 1974. “Judgment under Uncertainty: Heuristics and Biases.” *Science* 185 (4157): 1124–31.

Turner, J.H. & Machalek, R.S. (2018): *The New Evolutionary Sociology: Recent and Revitalized Theoretical and Methodological Approaches*, New York & London: Routledge.

Veblen, Thorstein. 1898. “Why Is Economics Not an Evolutoinary Science?” *Quarterly Journal of Economics* 12 (4): 373–97.

Voland, E. & Grammer, K. (Eds) (2003): *Evolutionary Aesthetics*, Berlin & Heidelberg: Springer.

Sociobiology Study Group of Science for the People (1976): ‘Sociobiology – Another biological determinism’, *BioScience* 26: 182-186.

Walsh, Denis. 2015. *Organisms, Agency, and Evolution*. Cambridge: Cambridge University Press.

Wilson, E.O. (1975a): *Sociobiology: The New Synthesis*, Cambridge, MA: Harvard University Press.

Wilson, E.O. (1975b): ‘For sociobiology’, *The New York Review of Books 22* (11 December 1975): 60-61.

Wilson, E.O. (1976): ‘Academic vigilantism and the political significance of sociobiology’, *BioScience* 26: 183-190.

Wilson, D.S. & Wilson, E.O. (2007): ‘Rethinking the theoretical foundations of sociobiology’, *Quarterly Review of Biology* 82: 327-348.

Witt U. (2003) *The Evolving Economy: Essays on the Evolutionary Approach to Economics*, Cheltenham: Edward Elgar.

Witt, U. & Chai, A. (Eds) (2019): *Understanding Economic Change: Advances in Evolutionary Economics*, Cambridge: Cambridge University Press.

Zebulum, R.S., Pacheco, M.A.C. & Vellasco, M.M.B.R. (2002): *Evolutionary Electronics: Automatic Design of Electronic Circuits and Systems by Genetic Algorithms*, Boca Raton, FL: CRC Press.

Ziman, J. (Ed.) (2000): *Technological Innovation as an Evolutionary Process*, Cambridge: Cambridge University Press.

Zurek, Wojciech Hubert. 2018. “Quantum Theory of the Classical: Quantum Jumps, Born’s Rule and Objective Classical Reality via Quantum Darwinism.” *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 376 (2123): 20180107.

Part I: How can disciplines benefit from, or contribute to, evolutionary frameworks?

Chapter 2

Is a non-evolutionary psychology possible?

Daniel Nettle[i], Thom Scott-Phillips[ii]

Abstract

The last thirty years has seen the emergence of a self-styled 'evolutionary' paradigm within psychology (henceforth, EP). EP is often presented and critiqued as a distinctive, contentious paradigm, to be contrasted with other accounts of human psychology. However, little attention has been paid to the sense in which those other accounts are not also evolutionary. We outline the core commitments of canonical EP. These are, from least distinctive to most: mechanism, interactionism, functionalism, adaptationism, and functional specialization. We argue that the minimal requirement for an approach to psychology to qualify as evolutionary in an important sense is functionalism. This is because the notion of functional design in organismal structures presupposes a history of evolution by natural selection. On this criterion, we argue, most, perhaps all of psychology qualifies as evolutionary, either implicitly or explicitly. We review several approaches that are typically contrasted with EP, showing that these are all evolutionary too by our criteria. We suggest that the EP/non-EP dichotomy be retired. However, though all psychology is evolutionary, psychologists do not always need to foreground evolution in their research, just as is true for biologists. At the same time, more space for evolution does not mean any less space for environment, context, culture, meaning or agency.

---

[i] Population Health Sciences Institute, Newcastle University, Department of Cognitive Sciences, Central European University.

[ii] Department of Cognitive Sciences, Central European University.

2.1. Introduction

The last thirty years has seen the emergence of a self-styled ‘evolutionary’ paradigm within psychology (henceforth, EP; Confer et al. 2010; Cosmides and Tooby 1987; Lewis et al. 2017; Tooby and Cosmides 1989). Whilst EP has been somewhat successful—judging by the appearance of textbooks, conferences and dedicated journals—it also continues to attract scepticism and critique. The claims of EP are often compared either unfavourably (by critics) or favourably (by adherents) to those of alternative paradigms in particular domains. These debates are sometimes presented, mistakenly as we shall argue, as pitting the merits of the ‘evolutionary perspective’ or ‘evolutionary thinking in psychology’ against those of some other perspective or type of thinking. However, it is not clear in what sense those other perspectives or types of thinking are not evolutionary too. We assume almost all scientists are committed to naturalism and accept that humans are organic creatures phylogenetically connected to other animals. Given these commitments, what it would mean for a psychology to be non-evolutionary is obscure. Would it suffice for the authors to not explicitly mention evolutionary principles, whilst not actually denying that humans are evolved creatures either, or is a non-evolutionary psychology something stronger than that?

In this chapter, we develop an account of what an account of psychology would have to look like to qualify as non-evolutionary, but we doubt that any actual productive exemplars exist. Though accounts of the mind vary enormously in detail, they tend to draw on a common set of ideas: that the mind contains mechanisms; that these are characteristic of humans and many are different from those of other species; and that they are loosely organized with respect to some notions of organismal function. These ideas are at least implicitly grounded in, and made coherent by, the fact that minds and brains are biological structures with evolutionary histories. Thus, a great variety of accounts of human psychology, including those drawn on by social scientists who would shudder to self-identify as evolutionary psychologists, are in some real sense evolutionary.

Canonical EP, the approach particularly associated with the work of Cosmides and Tooby (e.g. 1987) has more specific commitments than just appealing to evolution. We review these in detail below, but, briefly, canonical EP is strongly adaptationist; it takes a particular view of the granularity of adaptive problems; and, relatedly, sees psychological mechanisms as functionally specialized as a direct result of genetic evolution. These are important and distinctive commitments, but they are ancillary to taking an evolutionary perspective *per se*. We shall show that supposed alternatives to EP take different views on some of these ancillary commitments, particularly the one related to functional specialization. This does not, however, make them 'non-evolutionary'.

We find ourselves in constant danger of denotational confusion in this chapter. We seek to argue, in essence, that many accounts of psychology that are not EP are in fact EP, but that they differ from EP in important ways. To try to navigate through this thicket, where helpful we try to stick to the following terminological rules: we use *canonical EP* to refer to the approach of Cosmides and Tooby (Cosmides and Tooby 1987; Tooby and Cosmides 1989); *'non-EP'* with scare quotes to refer to approaches that, in discussions of the field, would commonly be considered non-evolutionary (erroneously, in our view); and *implicit EP* to refer to approaches to psychology that are meaningfully dependent on evolution, but do not draw attention to this fact. Hence, our central argument becomes the slightly more comprehensible claim that many 'non-EP' accounts of psychology are in fact implicit EP, but differ from canonical EP in important ways. The term *explicit EP*, and remaining uses of *EP* with no modifier, refer to work that self-identifies as evolutionary whether or not it exactly fits the canonical pattern.

In what follows, in section 2, we briefly sketch the commitments of canonical EP. In section 3, we raise and address a common misconception, which is that being 'evolutionary' in psychology leads to the expectation that human behaviour is inflexible, or controlled by genes, while being 'non-evolutionary' implies that humans are malleable and responsive to context. This has to be removed from the table before any serious analysis can begin. In section 4, we consider what would have to be true of a psychology for it to be non-evolutionary. We do this by undoing ever more of the commitments of canonical EP, until the point where all the evolution is gone. We conclude that a non-evolutionary psychology could exist in principle, though it would be strange and unattractive. Moreover, we see scant empirical evidence of its actual representation in the various approaches to psychology. In section 5, we examine some specific examples of 'non-EPs', or alternatives to canonical EP. We show that these are, implicitly and sometimes explicitly, evolutionary, and hence all qualify as broad sense EP. Section 6 concludes.

2.2. The commitments of evolutionary psychology

Canonical EP is founded on the commitments listed below. We review them only briefly as they have been very clearly stated elsewhere (Cosmides and Tooby 1987; Kurzban and Haselton 2006).

C1. *Mechanism.* Human cognitive contents and manifest behaviours are outputs of psychological mechanisms. These mechanisms have properties.

C2. *Interactionism.* Psychological mechanisms process environmental inputs in order to create cognitive contents and manifest behaviour. Because of this, neither cognitive contents nor manifest behaviours are invariant within or between humans. Patterns of environmental input explain variation in cognitive contents and manifest behaviours, but only do so satisfyingly when considered in conjunction with the characteristic ways that the psychological mechanisms detect, process and transform those inputs.

C3. *Functionalism*. Psychological mechanisms can be usefully viewed as exhibiting some kind of functionality for the whole organism. For example, in much the way that the immune system can be viewed as having the organismically-useful function of fighting infection, the visual perception system can be viewed as having the organismically-useful functions of detecting and recognizing things, and permitting movement around the environment.

C4. *Strong adaptationism.* Functional design arises from natural selection, and acknowledging this explicitly is epistemically useful. Because natural selection is a powerful force, we can assume, at least as an initial gambit, that the properties of many psychological mechanisms represent highly efficient design solutions to adaptive problems. This allows these mechanisms to be both ‘forward engineered’ (predicting as yet unobserved properties from a prior consideration of adaptive function) and ‘reverse engineered’ (inferring adaptive function from observed properties of the mechanism).

C5. *Functional specialization.* Psychological mechanisms are specialized in the functions they serve (for example, ‘detecting cheaters in exchange relationships ’or ‘choosing a suitable mate’, rather than ‘learning about the world’). This adaptive specialization drives the kinds of inputs that they respond to and the processing principles they employ. For a core set of human psychological mechanisms, this specialization is directly genetically specified.

Given these commitments, there is a sense in which the agenda of canonical EP was misnamed from the beginning. A more precise term would have been ‘Adaptationist Cognitive Science’; the ‘Evolutionary’ is potentially quite misleading. To see why, it is worth considering

Tinbergen's famous four-question typology for the explanation of behaviour (Ariew & Panchanathan, this volume; Tinbergen, 1963): 1. proximate causation; 2. ontogeny or development; 3. adaptive value or function; and 4. phylogeny or evolutionary history. Since it was only the fourth of these questions to which Tinbergen applied the term 'evolutionary', the reader might take that to be the one EP is addressing. Hence, the concern critics of canonical EP show for the paucity of verifiable information about behaviour of extinct hominins, and the lack of comparative evidence in many cases. However, this is a red herring, since canonical EP is not primarily concerned with Tinbergen's question 4.

If not question 4, readers might assume that canonical EP addresses Tinbergen's question 3, since this is a more obviously 'evolutionary' question than 1 and 2. In this case, since adaptive value or function is about the relationship of behaviour to fitness, canonical EP's lack of interest in counting babies or measuring life expectancies again seems anomalous. But again, this is a false friend. Although canonical EP uses theories or axioms concerning the relationships between behaviour and fitness, these relationships are not the explanatory target. Indeed, the assumptions made about them are generally uncontentious (in ancestrally-relevant human environments, there was gravity, objects were solid, there were males and females, animals and plants, social relationships, threats, allies, etc.). Canonical EP is actually concerned with question 1, figuring out how proximate psychological mechanisms work. Its distinctive approach to this task is to assume that explicit consideration of adaptive function will help guide the project of understanding proximate mechanism. That is, canonical EP claims that to progress in answering Tinbergen's question 1, it can be epistemically useful to bear in mind the possible answers to Tinbergen's question 2. Tinbergen's questions are distinct, but not completely irrelevant to one another. Nonetheless, the explanatory target for canonical EP is question 1, just as it is for any other kind of psychology.

Commitments C4 and C5 are worthy of some further comment. The scope of psychological mechanisms is *a priori* unknown. For example, is learning which foods are good to eat delivered by the same mechanisms as learning which people to trust, or are they delivered by two distinct mechanisms? Is figuring out how to build a fire the same kind of problem as figuring out how to comfort a friend, or are they both outputs of the capacity to reason? A very similar grain problem applies to adaptive problems: is avoiding predation the same adaptive problem as finding a mate, or are they both just instances of trying to maximize fitness (Sterelny and Griffiths 1999)? In the decades prior to the development of EP, evolutionary biologists had found it useful to carve the problem of maximizing fitness into a set of sub-problems that could be specified somewhat independently of one another. This led to the generation of a number of important middle-level models: parental investment theory (Trivers 1972), starvation-predation trade-off theory (Lima 1986), reciprocal altruism theory (Trivers 1971), and so on. Each of these models delineated a particular sub-part of the general problem of surviving and reproducing, showing that specific environmental or somatic factors would be especially relevant, and particular kinds of responsiveness to context might be favoured by selection in that domain.

Canonical EP made the gambit of assuming that the scope of psychological mechanisms corresponded one-to-one with to the scope of mid-level evolutionary biological theories (Cosmides and Tooby 1987). For example, there was a body of evolutionary theory concerning social exchange, the cooperation of unrelated individuals for mutual fitness benefit (Axelrod and Hamilton, 1981; Trivers 1971). Canonical EP assumed there would be psychological mechanisms specialized in making the computations required to implement such social exchange successfully: recognizing and remembering exchange partners, detecting cheating, calculating costs and benefits appropriately and so on (Cosmides and Tooby 1989). Now, there is no necessary reason this correspondence in scope has to hold: evolutionary biologists might not have carved up the problem space of fitness in the same way natural selection has carved up the computational space of the mind. The scope correspondence assumption was just a starting point: if it did not lead to novel predictions or insights about the mind, it could be revised, either globally or for a specific case. But, it was a distinctive programmatic offering for how one might begin the difficult business of studying psychological mechanisms by delineating the problem space in a way that was not totally arbitrary. In particular, it meant that mid-level theories from evolutionary biology became resources for generating testable hypotheses about what kinds of inputs psychological mechanisms would respond to and in which ways.

One consequence of the scope correspondence gambit is that the number of distinct proximal psychological explanations needed to account for the human mind is large: (at least) one for mate choice, one for social exchange, one for food selection, and so on. It is also necessitates some specification of how the otherwise distinct mechanisms that perform these specialized tasks are integrated together (the ‘architecture’ of the mind). This architectural project has been less thoroughly pursued within canonical EP than the project of trying to identify and study the individual specialized mechanisms (though see Barrett 2015; Sperber 2005; Tooby 2020). The assumption, because of C5, that there is a multiplicity of functionally distinct mechanisms stands in contrast to other approaches in psychology, where a small number of broad-grain explanations (associative learning, Bayesian updating, social learning strategies, the need for routine) is proffered as sufficient.

There are two more points to make in this section. First, much research in psychology effectively adopts all of C1-C5 without self-designating as EP. A great deal of research in cognitive development, for instance, aims to describe the functionally specialized cognitive processes that emerge in a robust and reliable way in typically developing children. As a specific example, when items are perceived as physical objects, this triggers in infants assumptions that the item is physically cohesive, bonded, rigid, and cannot be acted on at a distance (Spelke 1990). 'Core knowledge' such as this spans multiple domains extending well beyond object recognition, also including knowledge about human sociality (Carey 2009). Theorizing in cognitive development is strongly informed by assumptions about the potential function of cognitive processes (C3/C4), and there is a corresponding focus on specialization to fulfill that function (C5).

Second, it is possible for accounts of some phenomenon to each follow all of C1-C5, and yet end up with quite different substantive claims. For instance, individual performance in reasoning tasks is modulated by the mode of presentation, and there are competing EP explanations of this (Cosmides 1989; Sperber and Girotto 2002; Sperber, Cara, and Girotto 1995; Fiddick, Cosmides, and Tooby 2000). Both sides of this debate adopt all of C1-C5, but they differ in their claims of what specific cognitive processes generate the empirically observed phenomena. Thus, the relationship between 'adopting an EP perspective' and actual hypotheses about psychological mechanisms is definitely one-to-many. There can be multiple overtly EP accounts, even if they are all of the canonical EP flavour. Between them there is epistemic competition and a need for inference to the best explanation, on the usual kinds of grounds of plausibility, parsimony, consistency with a range evidence, generativity of novel predictions, and so forth. Thus, EP, even in its canonical form, is not strictly a theory. It is more like a meta-theory; or, even looser than that, a framework for how to set about generating and improving theories.

2.3. EP does not privilege genetic over environmental control of behaviour

As evolutionary psychologists have been at pains to point out (Al-shawaf et al. 2020; Cosmides and Tooby 1987; Kurzban and Haselton 2006), nothing in C1-C5 implies fixity or uniformity about human manifest behaviour or cognitive contents. On the contrary, mid-level theories from evolutionary biology show that the optimal strategy for solving an adaptive problem usually depends on the environment, on the individual's own somatic state, and the available alternatives. The point of having a psychology is to be able to process these contextual variables and respond appropriately to them. Thus, a rather general meta-expectation arising from EP is that evolved organisms, including humans, should be highly sensitive to context (by which is meant, broadly, features of the social and non-social environment), and also to their own somatic state. Individuals that persisted in doing or believing just the same thing regardless of the environment or their own state would have very poor fitness prospects indeed.

Thus, as a first-order generalization, in many cases we should expect manifest behaviour and cognitive contents to be changed by environmental inputs, broadly defined. Hence behaviour and cognitive contents should be different for individuals who have different experiences. However, as a second-order generalization, this responsiveness to context should be somewhat systematic. That is, the same combinations of environmental inputs should affect all humans in similar ways. For example, people the world over should tend not to be sexually attracted to others who grew up in close proximity or in association with the same mother, even if those others are otherwise compatible mates. There is considerable cross-cultural evidence for this and like claims. There is no tension between the claim and that facts that *who* grows up in close proximity varies a lot (for example, because of polygamy or institutions like kibbutzim), and that other features of the social environment can moderate the consequences of this tendency.

The most important claims of EP concerns not responsiveness to context (which it is compatible with), but the causal relevance of evolution and adaptation in explaining responsiveness to context. That is, humans are responsive to context in large part *because* they possess genetically evolved psychological mechanisms to detect those contextual features and respond to them in characteristic ways. Despite clear and repeated statements that EP is compatible with responsiveness to context (Al-shawaf et al. 2020; Cosmides and Tooby 1987;

Sperber and Hirschfeld 2007), EP continues to be held up as incompatible with responsiveness to context. For example, Levy writes 'if EP is correct, we should [predict] that our preferences remain stable across different environments' (Levy 2004, 461). Hence, EP is often seen as opposed to, and possibly refuted by, the basic insight of social science (the idea that context is what determines behaviour has been described as 'the central dogma of all social sciences', (Glass and Bilal 2016, 246). Though this misunderstanding has often been clarified, our informal impression is that it persists in the field. We briefly give several possible reasons that it does.

The first is that although EP is in principle compatible with responsiveness to context, some presentations describe cross-cultural recurrence of the same behaviours and ideas as the decisive evidence in EP's favour (Pinker 2002). Empirical EP studies have looked for universal patterns of sex differences (Schmitt and International Sexuality Description Project 2003), or universals in the content of moral systems (Curry, Mullins, and Whitehouse 2019). It is thus easy to infer from *tokens* that concern universals in manifest behaviour or cognitive contents that the EP *type* can only account for these, not for systematic variation. It is also true that many of the simple universal accounts inspired by EP are context-deaf and manifestly inadequate. For example, a simple EP account of the prevalence of obesity in the Western world might claim that humans, universally, have an unregulated appetite for sweet and fatty food, because these were scarce in ancestral environments, and hence there was no selection pressure for a regulatory mechanism to exist. This account fails: (a) because the existence of psychological mechanisms that regulate sugar consumption has been known about for decades (Cabanac, Duclaux, and Spector 1971; Fantino et al. 1983); and, more importantly, (b) it fails to account for the fact that obesity in the West mainly afflicts the poor, especially poor women; people who have the resources to choose just whatever foods they like are much more likely to be slim (McLaren 2007). These patterns are in fact compatible with a more sophisticated EP account, in which food choice and metabolic mechanisms take food availability and predictability as an important environmental input (Nettle, Andrews, and Bateson 2017). Nonetheless, the simple, context-deaf accounts have had sufficient airtime to get taken as representative of the potential of the type.

Second, it is very easy to mistake the claim that genetic evolution and adaptation are highly causally relevant to how manifest behaviour responds to environmental context, for the more direct claim that genetic evolution and adaptation are directly causally relevant to manifest behaviour itself. Tooby and Cosmides (1987) are very careful to distinguish between these two claims, but other commentators are not. For example, Smith (2020, 39) writes: ‘evolutionary psychologists believe that they have an inferential strategy that allows them to give accurate evolutionary explanations for contemporary human behavior’. Once this conflation has been made, it leads to an intuition of simple zero-sum competition for explanatory relevance: more for genes and adaptation means less for context and culture (for example ‘the debate between [standard social science] and EP concerns, not whether behaviour is the product of genes or the environment, but the relative importance of each’ (Levy 2004, 461)). This zero-sum view is a fallacy, since the explanatory target of EP (canonical EP at any rate) is not manifest behaviour, but the design features of psychological mechanisms, those design features including sensitivity to context. Nonetheless, it is an easy fallacy to fall prey to. Moreover, scholars often need to invoke straw figures to define the conceptual space in which they want to position their findings. The rather ubiquitous social science observation that context (or culture) matters has more impact when set against some kind of null hypothesis that context (or culture) should not matter, and EP gets cast, wrongly, in this role.

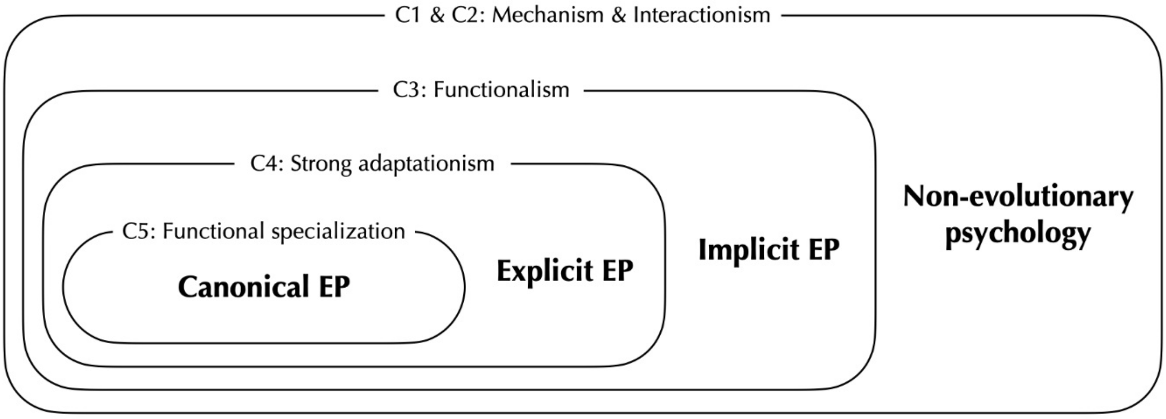


Figure 1: *Typology of possible psychologies based on which subsets of commitments C1 to C5 they adhere to. For explanation see text.*

2.4. What would constitute a non-evolutionary psychology?

In this section, we tackle the question of what the concept of a non-evolutionary psychology could possibly represent. We can think of commitments C1 to C5 as defining a series of nested subsets of accounts of psychology (figure 1). Any account committed to all of them would be in the same subset as canonical EP, and henceforth, by our logic, would be a form of canonical EP, even if it disagreed with the approach of Cosmides and Tooby in matters of detail. The question of this section is: how many of C5 to C1 have to be abandoned before the account is not evolutionary any more?

C5 could be relaxed without the approach becoming remotely non-evolutionary. We have already pointed out that canonical EP's correspondence assumption (one adaptive problem is subserved by one suite of mechanisms, and one suite of mechanisms solves one adaptive problem) is a gambit, a gambit that could turn out to be wrong. One could instead hypothesize, or discover, that evolution by natural selection has produced a small number of powerful, super-mechanisms whose computations help solve multiple adaptive problems (see e.g. Bolhuis, Brown, Richardson, & Laland, 2011). Examples might include associative learning, reasoning ability or social learning strategies (see section 5). We are not endorsing such an account: it is unclear that such devices, without further constraints, would be computationally adequate for achieving all the things that minds achieve (Sperber, 2005). Our point is that such an account could still be adaptationist. The super-adaptations, one could argue, have evolved precisely because they provide great adaptive flexibility in a relatively simple way, and do so highly efficiently. The term 'domain general' is often used for the mechanisms posited in such approaches, but it is somewhat misleading: mechanisms always have restrictions of domain. 'Generality' is at most a relative term. No-one seriously believes that the mechanisms involved when one acquires a suntan or immunity to a virus are the same ones with which one learns to dance. Thus, every mechanism has an input domain, defined by the kinds of representational content that can activate its processes, and the classes of operation it performs on that content. Given this, though, it makes sense to distinguish between more specialized mechanisms with narrower domains, and those whose domains are broader. Psychological mechanisms having broader domains than envisaged by canonical EP is a perfectly evolutionary proposal.

A more promising proposal is that a psychology would not be evolutionary if it were not committed to C4. As we suggested above, canonical EP might have been better named

‘Adaptationist Cognitive Science’, implying that adaptationism is a constitutive feature. Explicit appeal to the design-producing powers of natural selection, and to the procedures of forward- and reverse-engineering based on adaptive considerations, are deeply embedded in the EP literature. We concede that to qualify as an explicitly evolutionary approach in psychology, some reference to, or use of, adaptationist principles is required (though note that this does require evolutionary psychologists to claim that every mechanism is an adaptation; exaptations, by-products and constraints are also recognized within a broadly adaptationist framework (Buss et al. 1998)). Since much of the literature in psychology makes no such explicit appeal, one might be tempted to conclude that much psychology is ‘non-EP’.

However, we can distinguish here between explicit EP and implicit EP. The former specifically invokes C4. The latter does not do so. However, if the account makes any appeal to notions of function in respect of psychological mechanisms (i.e. C3), then we would still dub it implicitly evolutionary, because the ultimate source of functional design is evolution by natural selection. Why do organisms have visual perception systems that are functionally organized to deliver object recognition and safe spatial navigation? They do because ancestral organisms that could better recognize objects and navigate their environments were more likely to survive and reproduce. Without a history of natural selection, there is no good reason that organisms should be so conveniently equipped to achieve such purposes. A claim about organismal function presupposes, and relies on, a claim about evolution by natural selection, even if the authors do not make this plain. C3 is in effect a weaker and covert version of C4.

One possible objection here is to claim that functional organization of minds can stem from other sources than natural selection: for example, ontogenetic processes, or cultural transmission (see Wertz and Moya 2019). As we shall see in section 5, though, claims that other processes produce functional organization turn out to be built on a bedrock of assumed genetic adaptations that allow this to happen. That is, an organism can be organized by environmental inputs within its lifetime because it has particular kinds of plasticity. That plasticity is in turn the outcome of natural selection on genes: ancestors that were plastic in this way outcompeted those that were less plastic or plastic in different ways (Dickins and Rahman 2012; Nettle and Bateson 2015). Relatedly, some psychological mechanisms can become functionally organized in certain ways through training and practice, without a history of natural selection on that function. The ability to read is an obvious example. However, reading is a *derived* function (Sperber & Hirschfeld, 2007), made possible by scaffolding on a set of mechanisms such as language processing and visual object recognition. Those mechanisms in turn have *evolved* functions. It is the functional organization of the constituent mechanisms with respect to their evolved functions that makes the derived function possible: you couldn't learn to read unless you were already predisposed to parse human language and be good at detailed visual discriminations. Thus, we argue, grounding the idea that a psychological mechanism fulfills a function always leads to an invocation of the functionally organizing power of natural selection, either directly via evolved functions, or more indirectly via derived functions that depend on evolved functions. As such, we insist that any psychology making the slightest use of C3 is implicitly evolutionary.

The next possibility is that a psychology would count as non-evolutionary if it made no use of C3. A psychology would be non-evolutionary if posited that there are psychological mechanisms, but claimed that these mechanisms have no function and are not organized towards delivering any particular outcome for the organism. This seems to us to be right—this is what a non-evolutionary psychology would look like. We just doubt that any such approach to psychology actually exists, and can't see that it would be at all generative if it did. Every subfield of psychology is deeply and pervasively imbued with informal functional notions, whether it concerns the function of perceptual abilities, of memory, of concepts, of conformity, or of stereotypes. Thus, on our typology, most or even all of psychology is implicitly evolutionary.

At this point, psychologists might be tempted to distinguish between peripheral psychological processes and central ones (versions of this distinction are to be found in Fodor (1983) and dual-process accounts of cognition (E. R. Smith and DeCoster 2000)). Peripheral processes such as hunger, thirst and object recognition are obviously homologous to those in other animals; for those processes, a notion of function, and a link to evolutionary history, is uncontentious. Thus, no-one would have a problem with the notion that the study of these peripheral processes is implicitly an evolutionary study. However, central processes, such as reasoning, meaning-making, and identity, feel different. Perhaps those processes are somehow unshackled from any simple function. Hence, they must be studied in some non-evolutionary way. We think this argument is unproductive. Peripheral and central processes might feel phenomenologically different, but they are all just psychological processes delivered by brains. The chain that begins with sound waves being transduced by the ear and ends with the symbolic meaning of the call to prayer being recognized is unitary and continuous. Any attempt to split it into two parts of different kinds, one peripheral and evolved, the other central and non-evolved, causes many more difficulties than it solves (Pietraszewski & Wertz, 2021).

Moreover, even accounts of 'central' human mental life also draw on some tacit notion of organismal function when it suits them to do so. In the Freudian scheme, for example, the unconscious or id is the obviously animal-like part, fulfilling the function of getting mates and attacking rivals. However, the moralistic, human-specific superego is also conceived in functional terms. The superego functions to allow humans to live together with one another in harmony by providing internal controls on interpersonal behaviour (Freud 2002). A human with no superego would, through conflict, lose the benefits of sociality, and thus be worse off. This can be translated into a claim about the fitness of ancestral humans who had less developed superegos.

We could multiply examples. It is hard to give an account of any psychological process that does not somehow draw at least implicitly on some idea of the function of the mechanism for the purposes of the organism. They might differ widely with respect to which functions are supposed most relevant—being a competent member of a social group, making the world comprehensible through a coherent system of meaning or a predictable set of routines, or acquiring individual material benefits—but these are all in some sense organismal functions. Thus, the field of psychology does in practice lean on at least C3 (as well as C1 and C2, which we take to be respectively a minimal requirement for an enquiry to count as psychology, and a truism). Although a non-evolutionary psychology could exist in theory, we don't see any evidence of it in practice.

2.5. Alternatives to evolutionary psychology?

In this section, we briefly examine several paradigms that have been or could be considered competitors to EP, in an attempt to characterize which commitments they share with canonical EP, and where they differ. In the examples we discuss, the difference generally resides in C5, functional specialization. Our choice of alternative candidates is not meant to be exhaustive: our goal is rather to consider just a sufficiently diverse range of candidates, as a way to illustrate the general theses that any psychology can be considered at least implicitly evolutionary.

2.5.1. Social role theory

Social role theory (Eagly and Wood 2012) offers an account of the origins of sex differences in personality and social behaviour that is at least sometimes explicitly presented as an alternative to an EP account (e.g. ‘Sex differences in social behaviour: comparing social role theory and evolutionary psychology’ (Eagly 1997)). Social role theory accounts for sex differences in social behaviour in the following ways. First, society distributes individuals of the two sexes differently into economic roles. Exactly how it does this depends on the ecology – exactly what kinds of productive activities are required – but also on men’s greater size, physical strength, and speed, and women’s inherently greater involvement in pregnancy and lactation. The role specialization thus produced is an efficient division of labour given the economic need and the physical and reproductive sex differences. Second, individuals of both sexes observe the distribution of men and women across roles, and from this, develop gender role beliefs, such as that men are intrinsically more agentic and dominant, and women intrinsically more communal and caring. These gender role beliefs face two ways: into the self, and out toward others. Self-facing, they become internalized, and become standards toward which the individual self-regulates, generating in effect a self-fulfilling prophecy: women behave more communally in part because they believe women are more communal. Outward-facing, gender role beliefs lead people to negatively sanction, or at least not reward, role-atypical behaviour in other people. Adults also use gender role beliefs in the socialization of children, in part preparing them for the set of expectancies that they will in fact face as adults.

Eagly and Wood are explicit that their account has an ultimate evolutionary component (Eagly and Wood 2012). In particular, they discuss the adaptive basis of sex differences in physical traits, such as size, strength and ability to gestate and lactate. However, they see psychological sex differences (in personality or motivation) as not the direct result of adaptive evolution, but rather as indirect products of the division of labour, and the ability of humans to infer and internalise role beliefs. However, the ability of humans to infer and internalise role beliefs, and also the ability of humans to create an efficient division of labour, must themselves depend on psychological mechanisms (albeit, ones that do not themselves differ by sex). It would be possible to be adaptationist about those mechanisms too. For example, one could model the evolutionary dynamics of under what circumstances it is beneficial to accept and internalize a socially defined role. An individual able to do so would gain benefits of coordination and smooth, conflict free membership of groups, though possibly at the expense of accepting roles with lower fitness prospects than they could forge alone. Such questions have been explored, for example by reproductive skew theorists (Kokko and Johnstone 1999; Vehrencamp 1983). They find that voluntary submission to limiting roles (or subordinate positions) is adaptive under some conditions, but not under others. For example, outside options, costs of conflict, and relative power matter. The predictions of such models could nuance the account of social role psychology considerably: psychological mechanisms for role internalization should not be expected to produce complete, unconditional compliance. Instead, the mechanisms might be designed to respond (with role rejection or questioning) to certain kinds of contextual cues. Both the foundational observations of social role theory (that sex-typical behaviours change as the division of labour changes), and the political aspiration (to free women from oppressive gender role expectancies) would fit well with the 'vigilant role-internalizing' psychology that such evolutionary models would help understand (see also related discussion in §5.4).

Social role theory is already, by its own admission, evolutionary, and is adaptationist about physical sex differences. There is no reason that it could not also be adaptationist about its postulated role-internalising mechanisms. The real difference from the canonical EP accounts to which it gets contrasted is only the functional specialization assumed. For social role theory, the adaptive problem is something like 'figuring out how to socially coordinate given the local ecology', and the proposed solution is role-internalizing psychological mechanisms. This is distinct from the adaptive problems being obtaining resources from males (for women), and access to fertile women (for men), and the solution being evolved preferences that intrinsically differ by sex. Thus, even if social role theory is right, it is not a challenge to the broad project of evolutionary, even adaptationist, analysis.

2.5.2. Cultural evolutionary psychology and dual inheritance theory.

Cultural evolutionary psychology has been presented by Heyes (2018) as a competitor to canonical EP, though she also acknowledges the shared commitments of the two approaches. Heyes argues that key psychological mechanisms seen by many evolutionary psychologists as genetic adaptations (for example, theory of mind) are in fact created through patterns of teaching and cultural transmission. These mechanisms are thus not direct products of genetic evolution. Although this appears a strong challenge to canonical EP, the differences are perhaps less profound than might be imagined. First, canonical EP admits of psychological abilities that do not directly represent evolved adaptations, but rather, derived abilities built up through socialization practices. Reading was mentioned above as an obvious example. Such abilities must build on the mechanisms that are innately there: object recognition, plus the capacity to process language. Heyes concurs with this view, specifically describing a 'starter kit 'of evolved abilities, such as social motivations, attentional biases for faces and voices, associative learning, and capacities for inhibitory control and working memory. Heyes is apparently happy to be adaptationist about this starter kit. Thus, both Heyes and canonical EP agree that there is a set of evolved psychological adaptations, and then a set of derived abilities that are built up, through pedagogy and social transmission, scaffolding on the adaptations, in particular societies. They simply disagree about which abilities belong in the two sets. Both would put literacy in the derived set. Heyes but not canonical EP would put theory of mind in the derived set too (see Jacob and Scott-Phillips 2021 for discussion). These are substantive differences, but they do not make cultural evolutionary psychology either non-evolutionary or non-adaptationist.

Again, the key difference resides in the type of functional specialization committed to. For Heyes, the evolved adaptations, such as associative learning, have a broad domain. Concomitantly, the granularity of adaptationist analysis is broad too (the causally relevant adaptive problem is 'acquiring the best things to do in the local environment', rather than canonical EP's finer-grained parsing of adaptive problems). This broad parsing is similar to that of dual inheritance theory (Boyd & Richerson, 1985; Richerson & Boyd, 2005), a distinct tradition of cultural evolutionary thinking that predates Heyes and differs in some ways. The notable claim in this tradition is that humans have a second system of inheritance, culture, which allows human societies to achieve complex non-genetic adaptations to their environment, thus explaining human success in developing technologies, institutions and skills suitable for many environments. We will not review this approach in detail here. We note, however, that at its heart lies a claim that humans possess a set of genetically evolved adaptations for learning from others (these are known as social learning strategies; examples include conformist learning, prestige-biased learning, and so on). Dual inheritance theorists are explicitly adaptationist about the evolution of these mechanisms, viewing them as near-optimal solutions to figuring out what to do in variable environments (Boyd & Richerson, 1985; Henrich & McElreath, 2003).

Again, the difference from canonical EP is in the domain breadth of the mechanisms, and hence, relatedly, in the implied breadth of the causally relevant adaptive problem. Social learning strategies are used equipotentially to acquire a great variety of different cognitive contents, because they are beneficial overall. Indeed, this is one of the important claims of the approach, because it means that certain behaviours that are in fact not genetically adaptive (such as altruism) can 'hitch-hike', getting transmitted as a side-effect of the fitness benefits across other domains of learning socially (Boyd & Richerson, 2009; Henrich, 2004). Thus, the parsing of the adaptive problem, in dual inheritance theory, is 'figuring out good ways to live in the current environment', rather than narrow, separate parsing for each content domain that we find in canonical EP.

2.5.3. Bayesian cognitive science

Bayesian inference is a method of statistical inference used to calculate and update probabilities under uncertainty. In the past three decades, roughly concomitant with the emergence of canonical EP, Bayesian inference has been used to formally model many cognitive processes, from perception to planning, language comprehension to learning (Anderson 1991; Jones and Love 2011; Tenenbaum et al. 2011; Zednik and Jäkel 2016). As one example, consider communication and other forms of social interaction. Here, both production and comprehension are subject to uncertainty. On the production side: what behaviour will most likely trigger the intended inference in this audience? On the comprehension side: what inference could have been intended, given this behaviour? Bayesian inference provides the formal tools with which to approach such questions as inter-related planning problems over others' mental states (e.g. Goodman and Frank 2016; Ho et al. 2019; Shafto, Goodman, and Griffiths 2014).

The generality and flexibility of this approach has led some researchers to describe Bayesianism as an overarching framework for cognitive science: "a unifying framework for… perception, learning, reasoning, language comprehension and production, social cognition, action planning and motor control, as well as innumerable real-world tasks that require the integration of these capacities" (Chater et al. 2011, 194). Cognitive processes are modelled as problems of inference under uncertainty, with model specification (for example, prior structure, representational formats) tailored to specific problems. Optimal performance is often calculated as a benchmark for actual human performance, but in general Bayesian cognitive science prospers without explicit reference to the conceptual tools of EP. Does it therefore offer a potential non-evolutionary psychology?

On the contrary, Bayesian cognitive science adopts strong versions of at least four, and arguably all five of the commitments we described in §2. In particular, in practice it defaults to a great deal of narrow functional specialization (C5), because separate computational models are developed for each specific task, as necessary. Bayesian cognitive science also tends to be adaptationist (C4). Optimal performance is routinely used as a benchmark for actual human performance. The expectation that humans might perform near-optimally must, in the end, be grounded in some appeal to natural selection, either directly or through some process of learning or plasticity that is itself evolved. These analyses are thus adaptationist in practice, whether or not they make the connection to genetic evolution explicit. As such, Bayesian cognitive science is in fact complementary to canonical EP (Pietraszewski & Wertz, 2011). Whereas canonical EP aims to describe the specific adaptive tasks that human minds have been selected to solve, Bayesian cognitive science provides formal tools with which to describe possible computational solutions to such problems.

2.5.4. Giddens' social theory

Here, we extend our analysis, as a proof of principle, to a framework that comes from the social sciences, and as such would be seldom thought of evolutionary. This is the account of human psychology given by Anthony Giddens in *The Constitution of Society* (Giddens 1984). This is a work of sociology, not psychology. Giddens' purpose in sketching an outline of what he assumes about human psychology is to identify the set of human capacities or dispositions that make human social practices and social structures possible.

Giddens admits of a range of human motivations and tendencies. He sees people as having some (relatively unspecified) unconscious motivations, a general ability to build up practical knowledge of how to function in their environments, and also a discursive ability which allows them to articulate, reinterpret, communicate, challenge and discuss, the activities of their lives. Distinctively, Giddens claims that most of the time, the social practices we perform are only diffusely motivated. That is, humans have a need for predictability, and hence will acquire routines from their environment and reproduce them, because in so doing they create predictability in daily life. This is without those routines satisfying any more specific personal need. He attributes this to a psychological mechanism, the ‘basic security system’, that finds unpredictability costly and threatening (p. 41ff). He does concede that the desire for predictability is not the only human motivation. Thus, even though the need for routine may be key in explaining the stability of many arbitrary-seeming social practices, people also reshape, and sometimes reject, these in accordance with their purposes. More specifically, in ‘moments of crisis’, where routines are not doing their job, people bring their practices into the focus of discursive consciousness, and may bid to change them in fundamental and self-conscious ways.

The basic security system by no means exhausts human psychology, in Giddens’ view, but it is a distinctive claim and we focus on it here. The basic security system is a psychological mechanism seen as functional for the individual. Hence, it would be possible to be adaptationist about it: under what circumstances would a general tendency to acquire and reproduce routines be adaptive? Indeed, a specific adaptationist account of the adoption of arbitrary routine and convention has recently been provided by Theriault et al. (2021). Drawing on the large literature on ‘predictive processing’, they point out that the brain responds to differences between what it expects and what transpires. This prediction error is, *ex hypothesi*, metabolically costly, and learning can be seen as the process by which the organism minimises it, by revising its internal models, and/or by changing its environment. Theriault et al. (2021) argue that there are a wide variety of circumstances where, within tolerances, it is individually beneficial to follow conventions and routines, exactly because the metabolic cost of prediction error is thereby reduced. Roughly speaking, if I do what you expect, *your* behaviour in response is closer to what I expect, and both your and crucially *my* prediction error cost is lower. As they put it: “fluent social interactions are metabolically efficient social interactions” (p. 118). Since metabolic cost is a fitness-related quantity, managing it is fitness-enhancing, and hence this is an adaptationist account. According to Theriault et al. (2021), we should expect human societies to contain a substantial quantity of routine that is fairly arbitrary in its content and not directly motivated by any consideration other than achieving predictability itself, exactly as Giddens claims. Like Giddens, Theriault et al. (2021) do not claim that predictable routine is the *only* human motivation. Rather, people have a plethora of other motivations, but these need to be quite strong to overcome the metabolic cost and social dysfluency of breaking routine. They thus surface when conventions are very costly with respect to other goals, in something like Giddens’ moments of crisis (see also discussion of role-internalizing psychology in §5.1).

Our aim here has been to show that a social science account apparently light years from EP still assumes notions of organismal functionality, and thus is at least implicit EP. Indeed, the abilities postulated in this account can be subject to explicit adaptationist analysis, making it explicit EP without doing violence to the key claims. Again, a key difference from canonical EP resides in the functional specialization of the assumed mechanisms. For Giddens, and presumably for Theriault et al. (2021), the functional domain of the routinizing mechanisms is broad: prediction-error-minimizing routines will appear equipotentially across many domains of life, and be explained in the same way. This is contrast to the multiplicity of narrow adaptive problems, met by multiplicity of distinct functionally specialized mechanisms, and hence multiplicity of different proximal explanations, that we find in canonical EP.

2.6. Conclusions

We have hunted high and low for a real example of a non-evolutionary psychology, and not obviously found one. Accounts of human psychology generally rely on some claims about human psychological mechanisms, which they see characterize, even if inchoately, as functional with respect to some set of organismal goals. This notion of functionality presupposes, directly or indirectly, a history of natural selection. Thus, accounts of human psychology are at least implicitly evolutionary. Hence, we suggest that the distinction between EP and 'non-EP' should be retired. Theories in psychology differ in substantive ways, but the dichotomy 'evolutionary or not' is not especially useful. Every theory should be scrutinized based not on a distinction of type, but on the content of its character: what are the mechanisms implied; are they computationally adequate for the task; what are the grounds for positing them; what predictions does the theory make, and how well are those predictions supported?

Some readers may be alarmed to discover that they have been evolutionary psychologists all this time. We would like to reassure them with some conclusions. First, as we have argued, the class of evolutionary psychologies is a very broad one. Accepting that one is, after all, an evolutionary psychologist does not force one to accept the exact arguments of canonical EP, either in general or for any specific case. Second, although all psychology is evolutionary psychology, this doesn't mean that all psychologists have to talk about evolution all the time. Biology is often said to be a discipline unified by the theory of evolution (Dobzhansky 1973). Certainly, few biologists to our knowledge identify what they do as 'non-evolutionary biology'.

However, this does not mean that all biologists talk about evolution explicitly in their research programmes. Biology mostly consists in figuring out how biological mechanisms work. Much of this figuring out goes on through the usual cycle of proximal hypothesis and test, with the theory of evolution as no more than a background presence. There is no reason to think that psychology should be different. Explicitly adaptationist reasoning will prove useful at appropriate points, perhaps to differing extents in different research topics, but the admission that all psychology is evolutionary psychology does not mean that evolution should always be explicitly foregrounded. We would see greater acceptance of the importance of evolution as an enrichment—the conceptual resources of evolutionary biology provide an additional source of theory and hypothesis, without taking anything away—rather than a constraint on what kinds of concerns should be pursued.

Having said that, we believe that the explicit link to evolution and adaption is very useful for psychology. This is because of psychology's characteristic level of analysis, what David Marr called the computational level (Marr 1982). The computational level abstracts from much of the detail of physical implementation, to focus on what problems the mechanism solves and the processes by which it does this. For a psychological mechanism, the question of what the problem the system solves has a privileged conceptual connection to important organismal outcomes: staying alive, gaining information about the environment, avoiding threats, finding food, maintaining social relationships and so on. These in turn have a privileged connection to fitness and evolution. The moment one adopts the computational level of analysis, one necessarily confronts issues of function for the organism in a very obvious way. This means that the claim 'nothing in psychology makes sense except in the light of evolution' might be at least as true as the claim 'nothing in biology makes sense except in the light of evolution'. Similarly, the claim 'all psychology must be evolutionary psychology' has perhaps more force than the (also true) claim that 'all psychology must be quantum psychology'. You could not have a psychology inconsistent with the principles of quantum mechanics, but considerations of quantum mechanics obtrude much less directly when studying psychological mechanisms than considerations linked to evolutionary history do.

Some readers may still be concerned that explanatory layers of great value are necessarily lost once we admit the relevance of genetic evolution to the human mind. These layers might be (variously) the layer of meaning; the layer of agency; the layer of culture; or the layer of social structure or context. A full answer to this lingering concern is beyond the scope of this paper, but we believe it to be misplaced (see Boyer, 2018; Nettle, 2018; Sperber, 1996 for further discussion). There is no zero-sum competition between the importance of evolution and the importance of meaning, society, agency or environment (see section 3). Taking an evolutionary perspective is more than just compatible with believing these layers to be important. These layers arise from humans, with all their attributes and capacities, interacting and communicating in specific material, social and symbolic contexts over time. Thus, the layers can *only* be understood naturalistically by including enquiry into the evolved attributes and capacities that humans have as part of the overall project.

References

Al-shawaf, Laith, David M G Lewis, Yzar S Wehbe, and David M Buss. 2020. “Context, Environment and Learning in Evolution Psychology.” *Encyclopedia of Evolutionary Psychological Science*, no. January. https://doi.org/10.1007/978-3-319-16999-6.

Anderson, John R. 1991. “Is Human Cognition Adaptive?” *Behavioral and Brain Sciences* 14 (3): 471–85. https://doi.org/10.1017/S0140525X00070801.

Axelrod, Robert, and W D Hamilton,. 1981. “The Evolution of Cooperation.” *Science* 211: 1390–96.

Barrett, H C. 2015. *The Shape of Thought*. New York: Oxford University Press.

Boyd, R, and Peter J Richerson. 1985. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Boyd, Robert, and Peter J Richerson. 2009. “Culture and the Evolution of Human Cooperation.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 364 (1533): 3281–88. https://doi.org/10.1098/rstb.2009.0134.

Buss, David M., Martie G. Haselton, Todd K. Shackelford, April L. Bleske, and Jerome C. Wakefield. 1998. “Adaptations, Exaptations, and Spandrels.” *American Psychologist* 53 (5): 533–48. https://doi.org/10.1037/0003-066X.53.5.533.

Cabanac, M., R. Duclaux, and N. H. Spector. 1971. “Sensory Feedback in Regulation of Body Weight: Is There a Ponderostat?” *Nature* 229 (5280): 125–27. https://doi.org/10.1038/229125a0.

Carey, Susan. 2009. *The Origin of Concepts*. Oxford: Oxford University Press.

Chater, Nick, Noah Goodman, Thomas L. Griffiths, Charles Kemp, Mike Oaksford, and Joshua B. Tenenbaum. 2011. “The Imaginary Fundamentalists: The Unshocking Truth about Bayesian Cognitive Science.” *Behavioral and Brain Sciences* 34 (4): 194–96. https://doi.org/10.1017/S0140525X11000239.

Confer, Jaime C., Judith A. Easton, Diana S. Fleischman, Cari D. Goetz, David M. G. Lewis, Carin Perilloux, and David M. Buss. 2010. “Evolutionary Psychology: Controversies, Questions, Prospects, and Limitations.” *American Psychologist* 65 (2): 110–26. https://doi.org/10.1037/a0018413.

Cosmides, Leda. 1989. “The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task.” *Cognition* 31 (3): 187–276.

Cosmides, Leda, and John Tooby. 1987. “From Evolution to Behavior : Evolutionary Psychology as the Missing Link.” *The Latest on the Best Essays on Evolution and Optimality*, 277–306.

———. 1989. “Evolutionary Psychology and the Generation of Culture, Part II. Case Study: A Computational Theory of Social Exchange.” *Ethology and Sociobiology* 10: 51–97.

Curry, Oliver Scott, Daniel Austin Mullins, and Harvey Whitehouse. 2019. “Is It Good to Cooperate? Testing the Theory of Morality-as-Cooperation in 60 Societies.” *Current Anthropology*, no. May: 000–000. https://doi.org/10.1086/701478.

Dickins, Thomas E., and Qazi Rahman. 2012. “The Extended Evolutionary Synthesis and the Role of Soft Inheritance in Evolution.” *Proceedings of the Royal Society B: Biological Sciences* 279 (1740): 2913–21. https://doi.org/10.1098/rspb.2012.0273.

Dobzhansky, Theodosius. 1973. “Nothing in Biology Makes Sense except in the Light of Evolution.” *The American Biology Teacher* 35 (3): 125–29. https://doi.org/10.2307/4444260.

Eagly, Alice H. 1997. “Sex Differences in Social Behavior: Comparing Social Role Theory and Evolutionary Psychology.” *American Psychologist* 52: 1380–83. https://doi.org/10.1002/9781444390933.ch15.

Eagly, Alice H., and Wendy Wood. 2012. “Social Role Theory.” *Handbook of Theories of Social Psychology*, no. January 2012: 458–76. https://doi.org/10.4135/9781446249222.n49.

Fantino, M., F. Baigts, M. Cabanac, and M. Apfelbaum. 1983. “Effects of an Overfeeding Regimen—the Affective Component of the Sweet Sensation.” *Appetite* 4 (3): 155–64. https://doi.org/10.1016/S0195-6663(83)80028-2.

Fiddick, Laurence, Leda Cosmides, and John Tooby. 2000. “No Interpretation without Representation: The Role of Domain-Specific Representations and Inferences in the Wason Selection Task.” *Cognition* 77 (1): 1–79.

Fodor, J. 1983. *The Modularity of Mind: An Essay in Faculty Psychology*. Cambridge, MA: MIT Press.

Freud, Sigmund. 2002. *Civilization and Its Discontents*. London: Penguin.

Giddens, Anthony. 1984. *The Constitution of Society. Outline of the Theory of Structuration.* London: Polity Press.

Glass, Thomas A., and Usama Bilal. 2016. “Are Neighborhoods Causal? Complications Arising from the ?Stickiness? Of ZNA.” *Social Science and Medicine* 166: 244–53. https://doi.org/10.1016/j.socscimed.2016.01.001.

Goodman, Noah D., and Michael C. Frank. 2016. “Pragmatic Language Interpretation as Probabilistic Inference.” *Trends in Cognitive Sciences* 20 (11): 818–29. https://doi.org/10.1016/j.tics.2016.08.005.

Henrich, J. 2004. “Cultural Group Selection, Coevolutionary Processes and Large-Scale Cooperation.” *Journal of Economic Behavior & Organization* 53 (1): 3–35. https://doi.org/10.1016/S0167-2681(03)00094-5.

Henrich, Joseph, and Richard McElreath. 2003. “The Evolution of Cultural Evolution.” *Evolutionary Anthropology* 12: 123–35. https://doi.org/10.1002/evan.10110.

Heyes, C M. 2018. *Cognitive Gadgets: The Cultural Evolution of Thinking*. Cambridge, MA: Harvard University Press.

Ho, Mark K., Fiery Cushman, Michael L. Littman, and Joseph L. Austerweil. 2019. “People Teach with Rewards and Punishments as Communication, Not Reinforcements.”

Journal of Experimental Psychology: General 148 (3): 520–49. https://doi.org/10.1037/xge0000569.

Jacob, P, and T. C. Scott-Phillips. 2021. “Is Mindreading a Gadget?” *Synthese*.

Jones, Matt, and Bradley C. Love. 2011. “Bayesian Fundamentalism or Enlightenment? On the Explanatory Status and Theoretical Contributions of Bayesian Models of Cognition.” *Behavioral and Brain Sciences* 34 (4): 169–88. https://doi.org/10.1017/S0140525X10003134.

Kokko, Hanna, and Rufus A. Johnstone. 1999. “Social Queuing in Animal Societies: A Dynamic Model of Reproductive Skew.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266 (1419): 571–78. https://doi.org/10.1098/rspb.1999.0674.

Kurzban, Robert, and Martie G Haselton. 2006. “Making Hay out of Straw: Real and Imagined Controversies in Evolutionary Psychology.” In *Missing the Revolution*, edited by Jerome H. Barkow. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195130027.001.0001.

Levy, Neil. 2004. “Evolutionary Psychology, Human Universals, and the Standard Social Science Model.” *Biology & Philosophy* 19 (3): 459–72. https://doi.org/10.1023/B:BIPH.0000036111.64561.63.

Lewis, David M. G., Laith Al-Shawaf, Daniel Conroy-Beam, Kelly Asao, and David M. Buss. 2017. “Evolutionary Psychology: A How-to Guide.” *American Psychologist* 72 (4): 353–73. https://doi.org/10.1037/a0040409.

Lima, Steven L. 1986. “Predation Risk and Unpredictable Feeding Conditions : Determinants of Body Mass in Birds.” *Ecology* 67 (2): 377–85.

Marr, D. 1982. *Vision: A Computational Approach*. San Fransisco: Freeman & Co.

McLaren, Lindsay. 2007. “Socioeconomic Status and Obesity.” *Epidemiologic Reviews* 29 (1): 29–48. https://doi.org/10.1093/epirev/mxm001.

Nettle, Daniel, C Andrews, and M Bateson. 2017. “Food Insecurity as a Driver of Obesity in Humans: The Insurance Hypothesis.” *Behavioral and Brain Sciences* 40: e105–e105. https://doi.org/10.1017/S0140525X16000947.

Nettle, Daniel, and Melissa Bateson. 2015. “Adaptive Developmental Plasticity: What Is It, How Can We Recognize It and When Can It Evolve?” *Proceedings of the Royal Society B: Biological Sciences* 282 (1812): 20151005. https://doi.org/10.1098/rspb.2015.1005.

Pietraszewski, D, and Annie E. Wertz. 2021. “Why Evolutionary Psychology Should Abandon Modularity.” *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691621997113.

Pietraszewski, David, and Annie Wertz. 2011. “Reverse Engineering the Structure of Cognitive Mechanisms.” *The Behavioral and Brain Sciences* 34 (August): 209–10. https://doi.org/10.1017/S0140525X11000380.

Pinker, Steven. 2002. *The Blank Slate: The Modern Denial of Human Nature*. London: Penguin.

Richerson, Peter J, and Robert Boyd. 2005. *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Schmitt, David P., and International Sexuality Description Project. 2003. “Universal Sex Differences in the Desire for Sexual Variety: Tests from 52 Nations, 6 Continents, and 13 Islands.” *Journal of Personality and Social Psychology* 85 (1): 85–104. https://doi.org/10.1037/0022-3514.85.1.85.

Shafto, Patrick, Noah D. Goodman, and Thomas L. Griffiths. 2014. “A Rational Account of Pedagogical Reasoning: Teaching by, and Learning from, Examples.” *Cognitive Psychology* 71 (June): 55–89. https://doi.org/10.1016/j.cogpsych.2013.12.004.

Smith, E R, and J DeCoster. 2000. “Dual Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems.” *Personality and Social Psychology Review* 4: 108–31.

Smith, Subrena E. 2020. “Is Evolutionary Psychology Possible?” *Biological Theory* 15 (1): 39–49. https://doi.org/10.1007/s13752-019-00336-4.

Spelke, Elizabeth S. 1990. “Principles of Object Perception.” *Cognitive Science* 14: 29–56.

Sperber, Dan. 2005. “Modularity and Relevance: How Can a Massively Modular Mind by Flexible and Context-Sensitive?” In *The Innate Mind. Structure and Contents.*, edited by Peter Caruthers, Stephen Laurence, and Stephen Stich, 53–68. Oxford: Oxford University Press.

Sperber, Dan, Francesco Cara, and Vittorio Girotto. 1995. “Relevance Theory Explains the Selection Task.” *Cognition* 57 (1): 31–95.

Sperber, Dan, and Vittorio Girotto. 2002. “Use or Misuse of the Selection Task? Rejoinder to Fiddick, Cosmides, and Tooby.” *Cognition* 85 (3): 277–290.

Sperber, Dan, and Lawrence Hirschfeld. 2007. “Culture and Modularity.” *The Innate Mind: Culture and Cognition* 2: 149–164.

Sterelny, Kim, and Tom Griffiths. 1999. *Sex and Death: An Introduction to the Philosophy of Biology*. Chicago: University of Chicago Press.

Tenenbaum, J. B., C. Kemp, T. L. Griffiths, and N. D. Goodman. 2011. “How to Grow a Mind: Statistics, Structure, and Abstraction.” *Science* 331 (6022): 1279–85. https://doi.org/10.1126/science.1192788.

Theriault, Jordan E., Liane Young, and Lisa Feldman Barrett. 2021. “The Sense of Should: A Biologically-Based Framework for Modeling Social Pressure.” *Physics of Life Reviews* 36: 100–136. https://doi.org/10.1016/j.plrev.2020.01.004.

Tinbergen, N. 1963. “On Aims and Methods of Ethology.” *Zeitschrift Fur Tierpsychologie* 20: 410–33.

Tooby, John. 2020. “Evolutionary Psychology as the Crystalizing Core of a Unified Modern Social Science.” *Evolutionary Behavioral Sciences* 14 (4): 390–403. https://doi.org/10.1037/ebs0000250.

Tooby, John, and Leda Cosmides. 1989. “Evolutionary Psychology and the Generation of Culture, Part I.” *Ethology and Sociobiology* 10 (1–3): 29–49. https://doi.org/10.1016/0162-3095(89)90012-5.

Trivers, Robert. 1971. “The Evolution of Reciprocal Altruism.” *The Quarterly Review of Biology* 46: 35–57. https://doi.org/10.1086/406755.

———. 1972. “Parental Investment and Sexual Selection.” In *Sexual Selection and the Descent of Man, 1871-1971*, edited by B Campbell, 136–79. Chicago: Aldine.

Vehrencamp, Sandra L. 1983. “A Model for the Evolution of Despotic versus Egalitarian Societies.” *Animal Behaviour* 31 (3): 667–82. https://doi.org/10.1016/S0003-3472(83)80222-X.

Wertz, Annie E., and Cristina Moya. 2019. “Pathways to Cognitive Design.” *Behavioural Processes* 161: 73–86. https://doi.org/10.1016/j.beproc.2018.05.013.

Zednik, Carlos, and Frank Jäkel. 2016. “Bayesian Reverse-Engineering Considered as a Research Strategy for Cognitive Science.” *Synthese* 193 (12): 3951–85. https://doi.org/10.1007/s11229-016-1180-3.

Chapter 3

Evolutionary Economics and the Theory of Cultural Evolution

Ulrich Witt[i][1]

Abstract

---

[i] Max Planck Institute for the Science of Human History (formerly Max Planck Institute of Economics), Jena, Germany; ulrich.witt@evoecon.mpg.de

---

[1] I should like to thank André Ariew for very helpful comments on an earlier draft.

3.1. Introduction

Humans have been undertaking deliberate productive and consumptive activities since ancestral times. It is only over the last two or three centuries, however, that the variety and refinement of economic activities has massively increased (see Maddison 2001). The capacities and inclinations that humans inherit have not changed much since millennia. But the intensifying process of cultural (in the sense of human-made rather than natural) evolution has brought forth the deep-going transformations of the economy. One may therefore expect that a research branch running under the label "evolutionary economics" represents works that are connected to and contribute to the theory of cultural evolution rooted in the modern Darwinian synthesis. In other scientific disciplines the attribute "evolutionary" indeed indicates their "consilience" (Wilson 1998) with the unifying Darwinian frame (see, e.g., van den Bergh 2018 for a recent survey). Not so in the case of evolutionary economics. Research under this label differs from mainstream economics. But there are many different approaches and interpretations and an identification with the general Darwinian frame is not common (see Witt 2008).

Many contributions adopt the attribute "evolutionary" just to signal that they replace the rational actor heuristic based on methodological individualism by selection analogies based on population thinking. In other contributions the attribute only serves as a synonym for a special kind of economic dynamics: They focus on processes of economic change – especially those related to innovations, technical progress, and structural transformations of the economy – rather than the equilibrium states. (These states figure prominently in canonical economic theory.) Even more confusing, contributions that do view the economy through the looking glasses of the modern theory of evolution, i.e. do share the ontological premises of cultural evolution theory, rarely sail under the label evolutionary economics. Like the differences in theoretical approaches, there are differences in the economic topics which the various brands of evolutionary economics address.

Under these conditions, it is an open question what insights can be expected *for* evolutionary economics – as well as *from* evolutionary economics – in an exchange with cultural evolution theory. A characteristic of the explanatory domain of economics is that deliberation and the pursuit of individual interests shapes economic behavior much more so than many other forms of human behavior. This fact may invite reflections on the part of cultural evolution theory regarding the role that human agency plays in its domain – to which the economy belongs, after all. Such reflections may, conversely, help evolutionary economics to make progress with respect to its lack of hypotheses on the very aims of human intentionality. What motivations guide economic decision making? When there are commonly shared motivations, they can be expected to influence the innovative adaptations in the course of economic evolution. What common heritable and/or culturally contingent motivational forces (presumably themselves a result of evolution) are there? The disciplines contributing to the theory of cultural evolution may offer important information on this.

By human time scales, the shaping influence of natural selection is extremely slow. Economic and cultural adaptations more generally are much faster. Are inter-personally shared motivational forces the shaping agent filling the room that natural selection has left, particularly in most recent times? Are there cases in which the collective outcome of the agents choices – whether intended or unintended – point to evolutionary mismatches concerning the underlying motivations? Are the innovative adaptations in the economic domain biased in a certain direction as a result of human intentionality? Are this relevant questions also in other domains of cultural evolution such as technology or the sciences?

The present paper has two purposes. First, it offers in the following section a brief discussion of the research interests of a variety of approaches related to evolutionary economics. A question that deserves special attention here is how the different research strands position themselves with respect to the modern Darwinian theory of evolution in the sciences more generally and the corresponding theory of cultural evolution in particular. The second purpose is to stimulate in the subsequent section of the paper a discussion on possible gains from trade between evolutionary economics and the theory of cultural evolution. The central issue here is the role of human agency. Finally, some thoughts will be given to the problem of a bias in the evolution of the human economy and its potentially unsustainable consequences.

3.2. Evolutionary Economics and the Contested Relevance of the Darwinian Theory

Over more than a hundred years now Darwin’s theory of evolution has inspired some economists to call for a revision of economic theory. The main attempts to infuse evolutionary thinking into economics came in three waves. Each of the waves is characterized by a different view of what evolutionary economics can or should be. Works in earlier waves are rarely recognized by research in the later waves, and none of the waves succeeded in really making an impact on the mainstream canon in economics.

A first wave was launched by the now largely forgotten works of Thorsten Veblen. Strongly influenced by the ongoing *fin de siècle* Darwinian revolution in the sciences Veblen (1898) believed that Darwin’s theory should provide a foundation for theorizing about the human economy. For his conception of economic theory he introduced the term “evolutionary economics”. In his major works Veblen (1899, 1914) presented cultural-anthropological studies of the unfolding of the American capitalist society and its institutions. The critical attitude of the studies towards the contemporary social conditions in the U.S. attracted much attention, though the theoretical foundations of the criticisms leave much to be wanted (see Cordes 2005). An own school (the American institutionalists) emerged under Veblen’s influence. Yet, his vision of economics as an evolutionary science was not upheld for long in this school (see Hodgson 2004). It never gained ground in mainstream economic theory[1].

In present days, “evolutionary economics” has a different, albeit somewhat elusive, meaning. Many contributions cite Nelson and Winter’s (1982) interpretation of evolutionary economics that does not advocate a Darwinian approach. Instead, these authors suggest a

---

[1] Several decades later, Nobel memorial prize laureate Friedrich Hayek (1979, 1988) developed a very similar conception – now under the influence of sociobiology and with completely different political connotations. He outlined a theory of societal evolution aiming at a rectification of market liberalism (Beck 2018). While it had some influence on politics his theory had little impact on both evolutionary and mainstream economics.

synthesis of three quite different conceptual inputs. Only one of them motivates their use of the attribute “evolutionary”. This is the concept of “*economic* natural selection”. It is a loose analogy to natural selection models in evolutionary biology adapted to economic problems, particularly to problems related to industrial change, innovation-driven competition, and growth processes. The assumption is that, because of their bounded rationality, firm organizations have no practical way of finding the optimum. They rather make decisions deemed satisfactory on the basis of decision routines. Nelson and Winter here refer to the theory of bounded rationality – the second conceptual input in their synthesis[2]. Organizational routines can be adapted, though not by way of a fully rational response behavior, but again in a routine way by intervention of higher order routines. Ultimately, the entire set of routines of a firm is subjected to market selection. Either a firm thrives and grows with well-adapted routines or it declines and is driven out of the market jointly with its routines. Hence, in this interpretation it is not individual, optimal choice that shapes the market process, but the selection of the better fit decision routines[3].

The two mentioned inputs to Nelson and Winter’s synthesis ultimately serve to redefine, extend, and improve the third, and most momentous, theoretical element, namely Schumpeter’s (1912, 1942) theory of economic development. [4] Schumpeter (1912) had outlined a framework for explaining the cyclical growth of capitalist economies in which waves of entrepreneurial innovations played the central role. Innovative entrepreneurs introduce path-breaking

---

[2] The theory of bounded rationality goes back to the Carnegie school of organizational psychology in organization science (March and Simon 1958, Simon 1979).

[3] Nelson and Winter (1982) derive the implication of their interpretation for industrial change and economic growth by means of simulations analogously to a replicator dynamics. Metcalfe (1994) and (2002) later exploited an analogy to Fisher’s principles and Kimura’s theorem respectively to develop a rigorous model of the competitive market selection process.

[4] As they make clear: “… the term ‘neo-Schumpeterian’ would be as appropriate a designation for our entire approach as ‘evolutionary’. More precisely, it could reasonably be said that we are evolutionary theorists *for the sake* of being neo-Schumpeterians – that is, because evolutionary theories provide a workable approach to the problem of elaborating and formalizing the Schumpeterian idea view of capitalism as an engine of progressive change” (Nelson and Winter 1982, p. 39).

production techniques, new goods and services, new forms of tapping, organizing, and trading resources. Together with their imitating followers they outcompete existing businesses and thus transform the economy in a process of "creative destruction" (Schumpeter 1942).

With their revival of Schumpeter's agenda focusing on innovations, industrial dynamics, structural change, and economic growth Nelson and Winter succeeded in establishing a vibrant "neo-Schumpeterian" school in economics. Yet, the actual research practice in the school rarely makes use of their natural selection analogy (see, e.g., Hanusch and Pyka 2007, Fagerberg 2003, Dosi and Nelson 2018). This leaves the references to Nelson and Winter's concept of "evolutionary" economics in limbo. In fact, particularly the empirical works of the neo-Schumpeterians are often theoretically framed in a way not much different from that of mainstream industrial economics research on these topics (which is partly claiming a Schumpeterian lineage as well, see Aghion and Howitt 1992).

The little use that is actually made in the neo-Schumpeterian research of the natural selection analogy can be put in perspective with Schumpeter's explicit rejection of such an analogy as an explanatory tool[5]. There are reasons to believe that Schumpeter was inspired by the diffusionist school in the German-language social and cultural anthropology of his time (Kobayashi 2014). In opposition to Darwinism that school focused on the diffusion of novelty and its transformative influence on the "Kulturkreis" in which novelty had emerged[6]. A more recent continuation of the diffusionist tradition can be seen in works such as Rogers (1965), David (1993), Arthur (1994) which are frequently considered an own research strand in evolutionary

---

[5] Schumpeter (1912), Chap. 7. The English translation of the chapter (omitted from the English edition of 1934) is Schumpeter (2002). Andersen (2009) nonetheless takes pains to present Schumpeter himself as a forerunner of the economic natural selection conception.

[6] The school was founded by the German geographer Friedrich Ratzel. Leading proponents of the school were the ethnologists Leo Frobenius and Fritz Graebner whose works Schumpeter knew, see the quotes in Schumpeter (1955, Part IV, Chap. 3, Sec. 2b).

economics (see Metcalfe 1988, Stoneman 2007, Hanisch and Pyka 2007). In a neo-Schumpeterian spirit – but abstaining from any evolutionary rhetoric – these works focus on the (competitive) diffusion of innovations as a cause of economic change and growth.

It is only recently and a in different context that – in yet a new wave – some economists have again turned to the Darwinian theory of evolution as a frame for their research. This latest wave of evolutionary thought in economics comes in two variants. One of them takes up a development in evolutionary philosophy (if one may label it so) that postulates a few abstract principles as hallmark of evolutionary processes across all disciplinary domains. Inspiration comes from Campbell (1965) who distilled the principles of variation, selection, retention (or replication) from the Darwinian theory, from Dawkins' (1983) "Universal Darwinism", and from Hull's (1981) "interactor-replicator" scheme. With reference to these works Hodgson (2002), Aldrich et al. (2008), and others have argued that the principles of variation, selection, and retention should also guide the development of theories on the evolution of the economy.

It has not become entirely clear, though, what can emerge from such a theorizing. Is the aim to prove that the abstract patterns apply to a seemingly very different domain such as that of economics? If so, much of the theorizing can be expected to reproduce empirical inquiries into economic transformation processes just to recast them in terms of the new concepts and vocabulary. As Vromen (2018) has argued in his critique of Generalized Darwinism this would amount to Kitcher's (1985) "unification-as-explanation". This is an explanatory concept that should not be mixed up with causal explanations in the traditional sense. If, alternatively, the aim is to guide inquiries into, and develop causal explanations of, concrete transformation processes in the economy, it is not clear what the advantage of the "Generalized Darwinism" approach (i.e. of starting theorizing from just the three abstract principles) is.

For example, the economic historian Mokyr (2002, 2016, 2019) considers the evolution of useful knowledge the main driver of the unfolding of the modern economy and its growth. For conceptualizing that evolution he makes use of the three principles. The equivalent of the variation principle are innovative activities from which the huge historical variety and variability of useful knowledge results. Mokyr claims that this kind of knowledge creation is "super-fecund". Therefore not all newly generated knowledge can be preserved. The selection principle must apply. A particular technique may, for instance, by chance or deliberate choice be picked and implemented while other variants are not pursued further. Retention or replication finds its equivalent in the inter-generational transmission process. Mokyr argues that the historical process accords with the three principles. But his post hoc interpretation seems to be building on, rather than adding to, the already existing causal explanations of why and how useful knowledge in the economy has been unfolding (i.e. explanations that have not been, and need not be, derived from the three principles)[7].

The other variant in the latest wave of evolutionary thought in economics adopts the modern Darwinian theory as a frame for investigating a research topic that looms large also in several other social sciences and in sociobiology. Focus is on the contingencies of altruistic, pro-social, other-regarding individual behavior in social interactions. In fact, there are significant overlaps and cross references with works in evolutionary anthropology such as Henrich (2004.) From an ontological point of view the approach taken by this research direction therefore differs fundamentally from that of the neo-Schumpeterian analogy constructions (which is perhaps the reason for why the label "evolutionary economics" is usually avoided in these works). Accordingly, industrial dynamics and innovative economic growth – the pet themes of the neo-Schumpeterians – are not addressed, except in relation to the problem of sustainability of the human economy, e.g., in Safarzynska and van den Bergh (2010 and (Witt 2021).

Instead, the main topic is a critique and correction of the conventional view in economics of the individual agent as an unconditional, self-centered utility maximizer. Large scale economic experiments (see Henrich et al. 2004) have shown that, contrary to the assumptions of the economic mainstream, human social behavior ranges from altruism and cooperation to

---

[7] This may not be entirely surprising as even in evolutionary biology these abstract principles – by their very nature as post hoc generalizations – offer no specific insights into how a species or an ecosystem evolves. For a critical discussion see Buenstorf (2006), Levit et al. (2011), Vromen (2012). For another attempt to implement the approach of Generalized Darwinism in the economic domain see Hodgson and Knudsen (2010, Chapters 6 and 7).

pure self-interest and opportunism[8]. A more differentiated perspective has important implications for understanding how institutions regulate the allocation of economic resources and how these institutions evolve. As Ostrom (2010) has shown they are often not the result of rational political design, and they do not function without a critical share of pro-social attitudes in the population. Another major point raised by the more differentiated perspective is the question of how other-regarding, pro-social behavior and its evolution can be aligned with the rational choice assumption that dominates in economics.

The latter question seems to drive a substantial part of the research efforts so far (Gintis et al. 2005, Gintis 2007, Bowles and Gintis 2011). In the framework of rational-choice game theory (as opposed to evolutionary game theory) altruistic tendencies in human behavior can easily be explained under a special assumption: The players do not maximize their own payoff. alone, but choose a strategy that maximizes the weighted sum of the payoffs which they *and* the other players together obtain. While there is evidence for situations in which this assumption holds, experience teaches that this is far from regularly being the case. Humans can often be observed to engage in self-interested opportunism (accounting only for their own payoff) which may even result in free-riding and deceptive behavior. Such a behavior is able to exploit agents with unconditionally pro-social attitudes. How and when are other-regarding, pro-social preferences then able to prevail – if they emerge at all – against self-interest in a population?

For answering this question, several hypotheses have been proposed within the framework of rational choice explanations. Among them is the hypothesis of commonly shared information on the performance of individual players in interactions that disciplines self-seeking behavior through reputation effects (see, e.g., Binmore 2006). Another hypothesis is that of social ostracism practiced by pro-social individuals, i.e. the threat of punishing defectors (provided the threat is credible, see Fehr and Gächter 2002). More broadly seen it can be argued the

---

[8] Experimental economists have also made this point since long, see, e.g., Güth et al (1982), Hoffman et al. (1996).

institutional adaptations are necessary for suppressing violence as a means of self-seeking behavior. These adaptations can be observed to emerge historically through the formation of dominant coalitions of powerful individuals and families (see North et al. (2009).

3.3. Evolutionary Economics and the Theory of Cultural Evolution – Any Gains from Trade?

When putting on the looking glasses of the modern Darwinian theory of evolution the transformation of the human economy appears as part of the global evolution on this planet and cultural evolution in particular. It may be asked, therefore, what new insights the theory of cultural evolution and its Darwinian perspective can offer regarding the evolution of the human economy. Conversely, the question can be raised what the briefly outlined research in evolutionary economics, broadly defined, may imply for the understanding of cultural evolution.

In the case of the most recent wave of evolutionary research (mentioned last in the previous section) the mutual gains from trade can easily be identified. This research direction has benefitted from being able to join the debate on paradigmatic problems that had originally been outlined in works on cultural evolution theory. Conversely, in a close dialogue with evolutionary anthropology, the research has cast some new light on these problems. To wit: it has been shown that rational choice explanations are possible for the evolution of pro-social behavior and the conditions under which it is more or less likely to occur. These explanations may be seen as complements of explanations on the basis of group selection and multi-level selection hypotheses proposed in evolutionary biology (see Wilson 2004, 2016). Building on explanations of the latter kind Ostrom (2014) has outlined how elementary economic institutions evolve. While they are able to stabilize pro-social behavior the stability of these institutions is itself contingent on a sufficiently high frequency of that behavior within the population.

In the case of the neo-Schumpeterian research it is less obvious what the mutual gains from trade are or can be. Are there any insights that could be used to amend the theory of cultural evolution? One of the difficulties seems to be that the time period under investigation in this research is that of the exceptionally rapid transformations in most recent history. The bulk of the changes occur *within* a generation. In contrast, the major concepts developed in the theory of cultural evolution relate to the inter-generational co-evolution of cultural and genetic traits[9]. Can neo-Schumpeterian evolutionary economics perhaps beef up cultural evolution theory by providing tools and hypotheses for analyzing the intra-generational cultural transformation processes?

Many of the works implicitly or explicitly elaborate on teaching and imitation (or diffusion) processes. These processes form the core of the intra-generational transmission of production techniques and business practices. On one side there are theory-laden approaches harboring refined diffusion hypotheses that aim at empirical generalizations. A good example is the literature on industrial life cycles (e.g., Klepper 1997, Buenstorf and Klepper 2009). On the other side, many studies make use of historical case study methodology. They explore, for example, how learning and diffusion processes have unfolded in different industries (e.g., Fransman 1996, McKelvey 1996, Murmann 2003) or how transfers of scientific and technical knowledge into the productive sectors have been organized (e.g., Stephan 1996, Zellner 2003). These contributions seem worthwhile to consider as a repository of hypotheses and methodology supplementing cultural evolution theory in regard to the rapid transformation processes in modern societies.

---

[9] Concepts such as the "dual inheritance" hypothesis (Richerson and Boyd 2005) pertain to conditions when social practices, customs, techniques, etc. change over several generations and can therefore be treated as intra-generational inertia. The composition of the traits in a population changes when they are selectively transmitted between generations through imitation and teaching according to their relative advanage. But since the cultural traits are assumed to also affect survival and reproduction chances of the living generations, it follows that the inter-generational changes in the frequency distribution of cultural traits in a population depend on the traits' joint adaptive advantage. see also Brown and Richerson (2014).

Unlike the diffusion dynamics (based on teaching and imitation), the emergence of innovations is more difficult to explain, not least because of epistemic problems. For that reason, neo-Schumpeterian research rarely addresses the creative parts of the innovation processes. If at all, focus is instead on the circumstances under which more or less innovations occur. Various studies thus investigate the institutional and technological conditions that tend to foster or to inhibit innovativeness (e.g. Lundvall 1992, Scotchmer 2004, Fagerberg and Srholec 2008).

From a macro perspective, tinkering, exploring, trial and error learning, and deviations from the dictate of tradition have for most of history been more frequent in some domains of technology such as the military than in other domains such as production techniques. Only in historically recent times and largely restricted to economies in the Northern hemisphere, innovative productive activities have vastly increased. This fact invites the conjecture that there has been a macro shift in cultural evolution. In this vein Mokyr (2002, 2016) points to the changing collective perception of the human role in nature and the emancipation from religious myths and beliefs. He considers this a macro shift associated with the Enlightenment movement that opened the room for innovativeness at all levels of society and the economy.

While the conditions that encourage innovativeness change as the result of complex societal processes, the actual use made of any growing room is, of course, a result of individual choices. Hence, the individuals' motivation to search for and try new activities comes into focus and with it the role of human agency in cultural and economic evolution. Theory development in this direction leaves much to be wanted. Since the neo-Schumpeterian approach emphasizes the population level rather than individual behavior (see Metcalfe 2008) motivational hypotheses are a neglected topic. (For different reasons the same holds for mainstream economics. see Robson 2001.) While the disciplines that contribute to the theory of cultural evolution and that are rooted in biology do reflect on behavioral dispositions (see Brown and Richerson 2014) motivational hypotheses are in short supply in these disciplines as well. For the further development of evolutionary economics (and the clarification of the role of human agency in cultural evolution) more needs to be known about generic features of needs, drives, intentional goals striving, etc. that drive the adaptation processes at the individual level.

3.4. Innovative Adaptations Cause a Bias in Economic Evolution

Human behavior can be motivated in different ways. Some motivational forces such as needs and drives are innate and part of the evolved genetic endowment. They are “human universals” (2000) which means that they are inter-personally shared with the usual genetic variance. Innate needs and drives affect human intentionality and the aims on which it focuses. Deprived needs trigger a motivation to secure or improve their satisfaction[10]. Other motivational forces are the result of non-cognitive learning (conditioning learning, see, e.g., Leslie 1996, Chap. 2.4) and the human capacity of social cognitive learning, goal setting, and deliberate goal striving (see Bargh et al. 2010).

The two kinds of learning processes with their individual learning history are responsible for most of the diversity in the human kaleidoscope of idiosyncratic individual intentions. Nonetheless there are also widely shared aims due to similar cultural influences. For instance, similar “models of behavior” exert a conformity influence in the case of cognitive goal setting and goal striving, and similar culturally contingent reinforcement schemes do so in the case of conditioning learning.

Of these motivational forces not all are of course economically relevant. Economic relevance arises where the motivation induces a striving for control and command of resources. As everywhere in living nature, the pursuit of such motivations faces constraints caused by the scarcity of existing resources relative to the multiple claims made in relation to them. Competition for the scarce resources, whether peaceful or violent, leads to the satisfaction of

---

[10] For a closer discussion of the role of innate physical and mental needs such as those for water, sleep, food, physical activity, sex, shelter, affection, care, positive self-image, social recognition and status, cognitive and sensory stimulation see Witt (2018).

some of the rivaling claims and the privation of others. It is useful to distinguish in this context between two kinds of resources, namely human resources (labor) on the one side and natural resources on the other. Hunter and gather technology and agriculture – both progressing extremely slowly – have for most of human history limited the overall amount of natural resources available per capita for human consumption. Struggles for control and command of resources therefore took the form of zero-sum games: Minor fractions of society instrumentalized innovations in weapon technology and governance practices to increase their control of human resources as well as their share in the available natural resources at the expense of the rest of society.

However, when cultural and political conditions are favorable to innovativeness, human intelligence and creativity allows innovative adaptations, i.e. ways to expand existing resource constraints by invention and discovery. This was especially the case when – most likely because of the conjectured macro shift in cultural evolution – fundamental institutional changes came about in the economies of the Northern hemisphere over the past few centuries (see Dudley 1991). As a consequence, the prospects of moving from zero-sum games to positive-sum games by better exploiting natural resources became a strong motivational force driving innovative economic activities (where they were not just motivated by pure curiosity).

The innovative adaptation process that was triggered allowed the human kind to gain an advantage in terms of a population growth. When, more recently, political power constellations changed in a more egalitarian direction, individual life expectancy also grew and economic prosperity increased (Rosenberg and Birdzell 1986, Fagerberg et al. 2007). With a slowly decreasing average life working time the exploitation of human labor decreased. The opposite happened with respect to the per capita command of natural resources. Innovative exploration methods and revolutionized industrial production technologies increased the availability of materials and, in particular, fossil energy[11]. No wonder, thus, innovations have been hailed as “Prometheus unbound” (Landes 1969) or “lever of riches” (Mokyr 1990). They are broadly called for and encouraged now in politics particularly in democracies not least because they have proved to moderate conflicts over the distribution of resource in society.

Thus, driven by innovative adaptations, a bias in the evolution of the economy has developed: With the accelerating pace of innovations, the share of natural resources on the planet that is used for anthropogenic purposes grows[12]. Yet, there is a prerequisite for the persistence of this bias, namely that the striving for the command of resources does not vanish as ever more resources indeed become available per capita. More in particular, do the motivations – economically relevant innate needs, learned wants, and cognitively set goals – that directly or indirectly drive innovative activities persist independent of the amount of natural resources already used for human consumption[13]? This is foremost a question of whether and when satiation effects occur at some level of prosperity (Witt 2017).

Concerning the innate needs there are obviously some for which a satiation level does exist, i.e. where the demand (per unit of time) for natural as well as human resources serving the need no longer increases. An example is the case of the need for nutrition. Humans have been on the edge of deprivation for most of their history. In contrast, with the income level that has been

---

[11] Energy is necessary for processing and transformation of materials, transport and, not least, fertilizer production. The massive tapping of fossil energy allowed scaling up industrial production by orders of magnitude while reducing the costs per unit of output. The resulting expansion of the resource constraints was and is reflected in the growing per capita real income, see Ayres and Warr (2009). The latter ultimately means that for every hour of human work a growing amount of natural resource became available for human consumption.

[12] The present argument of a bias should not be mixed up with the idea of a directed evolution in a teleological sense. The evolution of the economy is not destined to progress towards superior stages as it was often postulated before the Darwinian revolution. An example is Herbert Spencer’s hypothesis of a progressing perfection of societal governance systems. For a sophisticated test – and rejection – of that hypothesis see Currie et al. (2010).

[13] Indirectly driven innovative activities are mediated by market exchange. They occur when producers anticipate needs, wants, or goals of their customers and aim to serve them by means of developing new goods and services.

reached in the majority of the contemporary economies this need is now easily satiable (Kaus 2012). There are several innate needs though that, for different reasons, are difficult to satiate or not at all satiable. A prominent example is the need for social recognition and status. As explained by Frank (2011), more and more resources are consumed in status races fueled by consumer innovations. Similarly, the need for cognitive and sensory stimulation causes a growing resource use. This is witnessed by the massively growing resource consumption of industries such as the leisure, entertainment, and tourism industry serving that need with their innovations[14].

Concerning learned wants and cognitively set goals there is no inherent boundary that prevents the emergence of ever new ones that motivate a striving for a greater resource command. In fact, producers in all industries with increasingly satiated markets are keen to launch innovations so as to offer new learning and goal setting opportunities and to propagate the new opportunities with intensifying commercial promotion activities. The purpose is to keep up the expansion of their sales. In effect the result is not a reduced, but a growing, use of natural resources for serving these motivations[15].

Accordingly, the bias in economic evolution towards an ever larger anthropogenic use of natural resources can be assumed to persist. But this growing use is not sustainable. Until the first warning signals by the "Club of Rome" (Meadows et al. 1972), natural resources had generally been perceived as nearly inexhaustible and only waiting for exploitation. Now it is

---

[14] See Lebergott (1993) for the statistical record of these industries in the U.S. Deeper reasons for the growth of these industries are discussed in Scitovsky (1976).

[15] Because of so-called rebound effects this holds for manufacturing industries despite the fact that the amount of natural resource used per unit of output decreases (Miklos and van den Bergh (2014). The implication also holds in spite of the fact that the service sector of the economy – less reliant on natural resources than industrial manufacturing – continually increases its value added share. The reason is that the relative growth of the value added share of the service sector is largely a statistical artifact. It results from the fact that the price level for labor-intensive services rises more quickly than the price level of resource-intensive manufactured goods (Henriques and Kander 2010).

successively discovered how relatively limited they are. In fact, in the case of some natural resources the anthropogenic use has already approached its "planetary boundaries" (see Steffens et al. 2015). The rapid progress of "Prometheus unbound" in making ever more natural resources available and – a concomitant of this progress – the strong growth of the size of the population with ever more humans laying claims to these resources, is on the way to threaten the ecosystem and its resilience.

The bias in the evolution of the economy thus tends to undermine the basis on which the human species makes a living. For that reason, it can be argued that a point has been reached where the combined effect of inherited and culturally contingent motivational forces and the human innovative capabilities reveals an evolutionary mismatch. The looming risks for the living and reproduction conditions of the human species seem to slowly be recognized in public. But while the sustainability problem now gains increasing attention, a solution for the motivational mismatch that drives the bias is far from being in sight.

3.5. Conclusions

This paper has briefly portrayed several contributions in economics that are related in one way or other to "evolutionary" theorizing. Some of these contributions use the attribute "evolutionary" in rather unspecific ways. Cases in point are inquiries into market competition, technological progress, and economic growth based on loose analogies to selection processes in nature. In other cases the attribute is just a synonym for the focus on a special, innovative kind of industrial dynamics. Still other approaches do adopt a Darwinian perspective in analyzing selected problems related to social behavior in economic interactions. The latter approaches have considerable overlap with topics that ae also on the agenda of cultural evolution theory. In fact, they have been developed in a dialogue with evolutionary anthropologists. Such an obvious overlap does not exist in the case of the less specific "evolutionary" approaches. Nonetheless they have something to offer for the theory of cultural evolution with their empirical studies and empirical generalizations focusing on the innovative adaptation process and the corresponding diffusion dynamics. They can serve as a repository of hypotheses and methodology supplementing cultural evolution theory in regard to the rapid transformations in modern economies.

As rapid as the intra-generational adaptations happen they are not shaped by natural selection forces. The question has therefore been raised especially with respect to the evolution of the economy what shaping influences govern these adaptations. The suggested answer starts from the fact that deliberation and the pursuit of individual interests strongly shapes economic behavior – more so than many other forms of human behavior. The course of economic evolution can be expected to reflect this fact. If there are commonly shared motivations and intentions they are likely to drive the innovative adaptations which the unique human intelligence and creativity makes possible. Since economics lacks own hypotheses on motivational commonalities it has been argued that evolutionary economics can benefit from an exchange with the disciplines contributing to cultural evolution theory that offer such hypotheses.

As the short discussion of corresponding hypotheses has shown, there are several commonly shared motivational forces that drive the innovative adaptation process in the economy. However, the economy may not be the only domain in which these shared motivations drive cultural adaptations. The role of human agency may therefore be worthwhile to scrutinize more generally in the theory of cultural evolution. As far as the role of human intentionality for economic evolution is concerned, it has been argued that, as everywhere in living nature, many shared motivational forces result in a striving for control of scarce resources. Yet, in combination with the unique human intelligence and creativity the striving causes a persistent bias in the intra-generational adaptations and, hence, in the evolution of the economy: the successful pursuit of the underlying motivations by means of innovative adaptations increases the dominance of the human species in the ecosystem. As a consequence, ever more resources of that system are claimed for anthropogenic uses. Once the growing claims on nature's resources start to threaten the ecosystem and its resilience, the bias in economic evolution reveals an evolutionary mismatch. This point seems to have been reached now. The inherited and culturally contingent motivational forces causing the bias then start to undermine the basis on which the human species makes a living.

References

Aldrich, H.E., Hodgson, G.M., Hull, D.L., Knudsen, T., Mokyr, J., Vanberg, V.J. (2008) In Defense of Generalized Darwinism, *Journal of Evolutionary Economics* 18, 577-596.

Andersen, E.S. (2009) *Schumpeter's Evolutionary Economics*. London: Anthem.

Arthur,W.B. (1994) *Increasing Returns and Path Dependence in the Economy*. Ann Arbor: University of Michigan Press.

Ayres, R.U. and Warr, B. (2009) *The Economic Growth Engine – How Energy and Work Drive Material Prosperity*. Cheltenham: Edward Elgar.

Bargh, J.A., Gollwitzer, P.M., and Oettingen, G. (2010) Motivation, in: S. T. Fiske, D. T. Gilbert, and G. Lindzey (eds.), *Handbook of Social Psychology*. New York: Wiley, 268-316.

Beck, N. (2018) *Hayek and the Evolution of Capitalism,* Chicago: University of Chicago Press.

Binmore, K. (2006) The Origins of Fair Play, *Papers on Economics and Evolution,* #0614, Max Planck Institute of Economics Jena.

Bowler, P.J. (1989) *Evolution - the History of an Idea*, Berkeley: University of California Press.

Bowles, S. and Gintis, H. (2011) *A Cooperative Species: Human Reciprocity and Its Evolution*, Princeton: Princeton University Press.

Brown, D.E. (2000) Human Universals and Their Implications. In: Roughley, N. (ed.) *Being Humans: Anthropological Universality and Particularity in Transdisciplinary Perspective*. New York: de Gruiter, 156- 174.

Brown, G.R. and Richerson, P.J. (2014) Applying Evolutionary Theory to Human Behavior: Past Differences and Current Debates, *Journal of Bioeconomics*, 16, 105-128.

Buenstorf, G. (2006) How Useful Is Generalized Darwinism as a Framework to Study Competition and Industrial Evolution? *Journal of Evolutionary Economics* 16, 511-527.

Buenstorf, G. and Klepper S. (2009) Heritage and Agglomeration: the Acron Tyre Cluster Revisited, *Economic Journal*, 119, 705-733.

Campbell, D.T. (1965) Variation and Selective Retention in Socio-cultural Evolution, in: Barringer, H.R., Blankenstein, G.I., and Mack, R.W. (eds.). *Social Change in*

Developing Areas: A Reinterpretation of Evolutionary Theory. Cambridge, Mass.: Schenkman, 19- 49.

Cordes, C. (2005) Veblen's 'Instinct of Workmanship,' its Cognitive Foundations, and Some Implications for Economic Theory. *Journal of Economic Issues* 39, 1-20.

Currie, T.E., Greenhill, S.J., Gray, R.D., Hasegawa, T., and Mace, R. (2010) Rise and Fall of Political Complexity in Island South-East Asia and the Pacific, *Nature*, 467, 801-804.

David, P.A. (1993) Path-Dependence and Predictability in Dynamical Systems with Local Network Externalities: A Paradigm for Historical Economics. In: D. Foray and C. Freeman (eds.). *Technology and the Wealth of Nations*, London: Pinter, 208-231.

Dawkins, R. (1983) Universal Darwinism. In: Bendall, D.S. (ed.). *Evolution from Molecules to Man*. Cambridge: Cambridge University Press, 403- 425.

Dosi, G. and Nelson, R.R. (2018) Technological Advance as an Evolutionary Process. In: R.R. Nelson et al. (eds.), Modern Evolutionary Economics, Cambridge: Cambridge University Press, 35-73.

Dudley, L.M. (1991) *The Word and the Sword – How Techniques of Information and Violence Have Shaped Our World,* Oxford: Blackwell.

Fagerberg, J. (2003) Schumpeter and the Revival of Evolutionary Economics: an Appraisal of the Literature. *Journal of Evolutionary Economics* 13, 125-159.

Fagerberg, J., Srholec, M. and Knell, M. (2007) The Competitiveness of Nations: Why Some Countries Prosper While Others Fall Behind. *World Development* 35, 1595-1620.

Fagerberg, J. and Srholec, M. (2008) National Innovation Systems, Capabilities, and Economic Development. *Research Policy* 37, 1417-1435.

Fehr, E. und Gächter, S. (2002) Altruistic Punishment in Humans. *Nature* 415, 137-140.

Frank, R.H. (2011) *The Darwin Economy*. New Haven: Princeton University Press.

Fransman, M. (1995) *Japan's Computer and Communications Industry - The Evolution of Industrial Giants and Global Competitiveness*, Oxford: Oxford University Press.

Gintis, H. (2007) A Framework for the Unification of the Behavioral Sciences, *Behavioral and Brain Sciences,* 30, 1-61.

Gintis, H., Bowles, S., Boyd, R., Fehr, E., eds. (2005) *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life,* Cambridge, MA: MIT Press.

Güth, W., Schmittberger, R. and Schwarze, B (1982) An Experimental Analysis of Ultimatum Bargaining, *Journal of Economic Behavior and Organization* 3, 367-388.

Hanusch, H. and Pyka, A., eds. (2007) *Elgar Companion to Neo-Schumpeterian Economics,* Cheltenham: Edward Elgar.

Hayek, F.A. (1979) *Law, Legislation and Liberty*; Vol. III: The Political Order of a Free People. London: Routledge.

Hayek, F.A. (1988) *The Fatal Conceit*, London: Routledge.

Henrich, J. (2004) Cultural Group Selection, Coevolutionary Processes and Large-scale Cooperation, *Journal of Economic Behavior and Organization,* 53, 3-35.

Henrich, J.P., Boyd, R., Bowles, S., Camerer, C., Fehr, E., and Gintis, H. (eds.) (2004) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-scale Societies*. Oxford: Oxford University Press.

Henriques, S. T. and Kander, A. (2010). The Modest Environmental Relief Resulting from the Transition to a Service Economy. *Ecological Economics* 70, 271–282.

Hodgson, G.M. (2002) Darwinism in Economics: From Analogy to Ontology, *Journal of Evolutionary Economics,* 12, 259-281.

Hodgson, G.M. (2004) *The Evolution of Institutional Economics*, London: Routledge.

Hodgson, G.M. and Knudsen, T. (2010) *Darwin's Conjecture – The Search for General Principles of Social and Economic Evolution*, Chicago: University of Chicago Press.

Hoffman, E., McCabe, K. and Smith, V.L. (1996) Social Distance and Other-regarding Behavior in Dictator Games. *American Economic Review*, 86, 653-660.

Hull, D.L. (1981) Units of Evolution: A Metaphysical Essay. In: U.J. Jensen, R. Harré (eds.), *The Philosophy of Evolution*, New York: St. Martin's Press, 23-44.

Kaus, W. (2012) Beyond Engel's Law – Pursuing an Engelian Approach to Welfare. A Cross Country Analysis, *Jena Economic Research Papers* 2012-28.

Kitcher, P. (1985) *Vaulting Ambition: Sociobiology and the Quest for Human Nature*, Cambridge, MA.: MIT Press

Klepper, S. (1997) Industry Life Cycles. *Industrial and Corporate Change* 6, 145-181.

Kobayashi, D. (2014) Effects of Anthropology and Archaeology Upon Early Innovation Studies, Paper presented at the International Schumpeter Society Conference, Jena 2014.

Lebergott, S. (1993) *Pursuing Happiness – American Consumers in the Twentieth Century.* Princeton: Princeton University Press.

Leslie, J.C. (1996) *Principles of Behavioral Analysis*, Amsterdam: Harwood Academic Publishers.

Levit, G.S., Hossfeld, U. and Witt, U. (2011) Can Darwinism Be 'Generalized' and of What Use Would This Be? *Journal of Evolutionary Economics* 21, 545-562.

Lundvall, B.-Å. (ed.) (1992) *National Systems of Innovation. Towards a Theory of Innovation and Interactive Learning*. London: Pinter Publishers.

Maddison, A. (2001) *The World Economy: A Millenium Perspective, OECD,* Paris.

March, J.G. and Simon, H.A. (1958) *Organizations*, New York: Wiley.

McKelvey, M. (1996) *Evolutionary Innovations: The Business of Biotechnology*. Oxford: Oxford University Press.

Meadows, D.H., Meadows, D.L., Randers, J., Behrens III, W.W. (1972) *The Limits to Growth*, New York: Universe Books.

Metcalfe, J.S. (1988) The Diffusion of Innovations: An Interpretative Survey. In: Dosi,G., Freeman,C., Nelson,R., Silverberg,G., and Soete,L. (eds.). *Technical Change and Economic Theory*. London: Pinter, 560- 589.

Metcalfe, J.S. (1994) Competition, Fisher's Principle and Increasing Returns in the Selection Process", *Journal of Evolutionary Economics,* 4, 327-346.

Metcalfe, J.S. (2002) On the Optimality of the Competitive Process: Kimura's Theorem and Market Dynamics, *Journal of Bioeconomics*, Vol. 4, 109-133.

Metcalfe, J.S. (2008) Accounting for Economic Evolution: Fitness and the Population Method, *Journal of Bioeconomics*, 10, 23-50.

Miklós, A. and van den Bergh, C.J.M. (2014) Re-spending Rebound: A Macro-level Assessment for OECD Countries and Emerging Economies. *Energy Policy* 68, 585-590.

Mokyr, J. (1990) *The Lever of Riches - Technological Creativity and Economic Progress,* Oxford: Oxford University Press.

Mokyr, J. (2002) *The Gifts of Athena – Historical Origins of the Knowledge Economy,* Princeton: Princeton University Press.

Mokyr, J. (2016) *A Culture of Growth: Origins of the Modern Economy.* Princeton: Princeton University Press.

Mokyr, J. (2019) Science, Technology, and Knowledge: What Economic Historians Can Learn From an Evolutionary Approach. In: U. Witt, A. Chai (eds.), *Understanding Economic Change – Advances in Evolutionary Economics*, Cambridge: Cambridge University Press, 81-119.

Murmann, J.P. (2003) *Knowledge and Competitive Advantage - The Coevolution of Firms, Technology, and National Institutions*. Cambridge: Cambridge University Press.

Nelson, R.R. and Winter, S.G. (1982) *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.

North, D.C., Wallis. J.J. and Weingast, B. (2009) *Violence and Social Orders,* Cambridge: Cambridge University Press.

Ostrom, E. (2010) Beyond Markets and States: Polycentric Governance of Complex Economic Systems. *American Economic Review* 100(3), 641-672.

Ostrom, E. (2014) Do Institutions of Collective Action Evolve? *Journal of Bioeconomics* 16, 3-30.

Richerson, P.J. and Boyd, R. (2005) *Not by the Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Robson, A.J. (2001) The Biological Basis of Economic Behavior. *Journal of Economic Literature* 39, 11-33.

Rogers, E. M. (1995) *Diffusion of Innovations, 4. ed.* New York: The Free Press.

Rosenberg, N., Birdzell, L.E. jr. (1986) *How the West Grew Rich - The Economic Transformation of the Industrial World*. New York: Basic Books.

Safarzynska, K. and van den Bergh, J.C.J.M. (2010) Evolving Power and Environmental Policy: Explaining Institutional Change with Group Selection. *Ecological Economics* 69, 743-752.

Schumpeter, J.A. (1912) *Theorie der wirtschaftlichen Entwicklung*. Leipzig : Duncker & Humblot. English transl.: *Theory of Economic Development*, Cambridge, Mass.: Harvard University Press 1934.

Schumpeter, J.A. (1942) *Capitalism, Socialism and Democracy*, New York: Harper.

Schumpeter, J.A. (2002) The Economy as a Whole (English translation of chapter 7 of Schumpeter 1912). *Industry and Innovation* 9, 93-145.

Schumpeter, J.A. (1955) *History of Economics Analysis*, London: Allen & Unwin.

Scitovsky, T. (1976) *The Joyless Economy.* Oxford: Oxford University Press.

Scotchmer, S. (2004) *Innovation and Incentives*, Cambridge, MA.: MIT Press.

Simon, H.A. (1979) Rational Decision Making in Business Organizations, *American Economic Review,* 69, 493-513.

Stephan, P.E. (1996) The Economics of Science. *Journal of Economic Literature* 34, 1199-1235.

Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Riggs, R., Carpenter, S.R., de Vries, W., de Wit, C.A., Folke, C., Gerten, D. Heinke, J., Mace, G.M., Persson, L.M., Ramanathan, V., Reyers, B., Sörlin, S., (2015) Planetary Bounderies: Guiding Human Development on a Changing Planet. *Science* 347, 1259853.

Stoneman, P. (2007) Technological Diffusion: Aspects of Self-propagation as a neo-Schumpeterian Characteristic. In: H. Hanusch and A. Pyka, (eds.), *Elgar Companion to Neo-Schumpeterian Economics,* Cheltenham: Edward Elgar, 377-385.

van den Bergh, J.C.J.M. (2018) *Human Evolution – Beyond Biology and Culture.* Cambridge: Cambridge University Press.

Veblen, T. (1898) Why Is Economics Not an Evolutionary Science? *Quarterly Journal of Economics,* 12, 373-397.

Veblen, T.B. (1899) *The Theory of the Leisure Class - An Economic Study of Institutions*, New York: MacMillan.

Veblen, T.B. (1914) *The Instinct of Workmanship, and the State of the Industrial Arts*, New York: MacMillan.

Vromen, J.J. (2012) How to Make a Convincing Case for Darwinizing the Social Sciences. *Journal of Economic Methodology* 19, 77-88.

Vromen, J.J. (2019) Generalized Darwinismin Evolutionary Economics. In: U. Witt, A. Chai (eds.), *Understanding Economic Change – Advances in Evolutionary Economics*, Cambridge: Cambridge University Press, 120-154.

Wilson, D.S. (2004) The New Fable of the Bees: Multilevel Selection, Adaptive societies, and the Concept of Self-interest. *Evolutionary Psychology and Economic Theory 7*, 201-220.

Wilson, D.S. (2016) Two Meanings of Complex Adaptive Systems. In: Wilson, D.S. and Kirman, A. (eds.), *Complexity and Evolution – Toward a New Synthesis for Economics*. Cambridge, MA.: MIT Press, 31-46.

Wilson, E.O. (1998) *Consilience - The Unity of Knowledge*. New York: Alfred Knopf.

Witt, U. (2008) Evolutionary Economics, in: S.N. Durlauf, L.E. Blume (eds.), *The New Palgrave Dictionary of Economics*, 2nd edit. Vol. 3, New York: Palgrave MacMillan, 67-73.

Witt, U. (2017) The Evolution of Consumption and Its Welfare Effects, *Journal of Evolutionary Economics*, 27, 273-293.

Witt, U. (2018) Evolutionary Economics and Psychology. In: A. Lewis (ed.), *Cambridge Handbook of Psychology and Economic Behavior*, Cambridge: Cambridge University Press, 673-693.

Witt, U. (2021) Does Sustainability-promoting Policy Making Reduce Our Welfare? *Ecological Economics* 188, https://doi.org/10.1016/j.ecolecon.2021.107130

Zellner, C. (2003) The economic effects of basic research: Evidence for embodied knowledge transfer via scientists' migration. *Research Policy* 32, 1881-1895.

Chapter 4

Repetition without Replication: Notes Towards a Theory of Cultural Adaptation

Carsten Strathausen[i]

Abstract

This paper proposes a theoretical framework for the study of cultural adaptation. A subfield within the larger field of cultural evolution, cultural adaptation is defined as the purposeful remediation of cultural artefacts into different (artistic, social, historical) contexts. My proposal combines insights from two distinct fields, extended evolutionary studies in the sciences and cultural adaptation studies in the humanities. The collaboration between these fields, I argue, has been hampered by neo-Darwinian reductionism and the false meme-gene analogy, on the scientific side, and a long-standing bias against statistical-quantitative approaches to culture, on the humanities side. Instead of a hierarchical approach that looks for core-units of culture (i.e., memes), I propose a relationist approach inspired by the Extended Evolutionary Synthesis that analyzes the dynamic network of interrelated products, processes, and receptions by which artistic material is continually refitted into different forms for new audiences. The paper is divided into two parts. The first part introduces scientific theories of cultural evolution, while the second part switches perspectives to the humanist study of cultural adaptations in the field of Adaptation Studies. As my title suggest, I shall use a series of theses to advance my argument throughout the paper.

---

[i] University of Missouri in Columbia, Missouri.

I

4.1. Cultural evolution is a Lamarckian, not a (neo-)Darwinian process

I am using the parenthetical prefix (neo-) to emphasize that Darwin himself was certainly *not* a neo-Darwinian in today's sense. Darwin knew nothing about molecular genetics or the Mendelian laws of inheritance that have come to dominate evolutionary theory since the modern synthesis (MS) in the early 20th century. Throughout his life, Darwin remained an ardent Lamarckian, who believed that the phenotype's acquired traits over lifetime are inheritable from one generation to the next.[1] Starting with Weissmann's distinction of germ-cells and soma-cell (1893), the scientific consensus during much of the 20th century was that Lamarckism does not apply to biological evolution, because the code of genetic replication is unaffected by ontogenetic development. It follows that any acquired traits of a phenotype during lifetime cannot be inherited by their off-spring. According to MS, the flow of information in living organisms is always one-directional. It runs exclusively from genotype to phenotype, never the reverse, and it renders human ontogeny, the development of individuals over lifetime, irrelevant to the process of replication.

Over the last three decades, this central dogma of MS—namely that genetic inheritance is *not just one, but the only relevant system of inheritance in living beings*—has come increasingly under attack both within and beyond the life sciences (cf. Strathausen, 2017). Most importantly, the central dogma has proven untenable empirically. There is abundant biological research and scientific evidence that demonstrates the importance of other, non-genetic systems of inheritance both at the biological and the social level. At the biological level, these systems include cellular epigenetic inheritance systems that can influence gene activation/regulation at the phenotypic level and co-determine the structure of certain proteins that interact with DNA (cf. Jablonka and Lamb 2020, 22-7). At the social level, these systems include behavioral

---

[1] The common distinction between Darwinism and Larmarckism remains extremely fuzzy and relative, Jablonka and Lamb argue: "When a developmentally selected variation is transmitted to the next generation of organisms, the effect is Lamarckian at the individual level, even when the generation of the variation depends on entirely blind processes at the intra-organismal level. Whether a process is regarded as 'Darwinian' or 'Lamarckian' therefore depends on the level of organization that is analysed" (61).

inheritance through social learning and symbol-mediated inheritance through language and other media. Jablonka and Lamb also mention “soma-mediated inheritance” as another system of inheritance (besides genetic, epigenetic, behavioral, and symbolic systems). Situated in-between the biological and the social realm, soma-mediated inheritance is defined as “the transmission of variations through the physiological reconstruction of the conditions in which organisms live and develop” (Jablonka/Lamb 2020, 27). They provide the following example:

> Female pups of rat mothers who gave them a lot of licking and grooming grow up to be adventurous and not readily stressed; they, in turn, give their own pups a lot of maternal care, so the cycle continues. Conversely, pups that are given less care grow up to be fearful and easily stressed, and show poor parenting, which leads to the perpetuation of the fearful behaviour in their own offspring. The outcomes of the differences in parental care are not due to inherited gene differences: they are the same even if a foster mother is used, so no information about parenting is transmitted through the germline. The information that leads to altered brain physiology in their young is transmitted through the mothers’ behaviour, not through their gametes. (Jablonka and Lamb 27).

A core principle of the Extended Evolutionary Synthesis, Jablonka and Lamb argue, is that these various systems of inheritance interact with each other such that information flows back and forth between them. Inheritable information, in other words, does not run in just one direction from genotype to phenotype, as the Modern Synthesis (MS) falsely claims. Instead, Jablonka and Lamb argue,

The variations generated and transmitted through different inheritance systems act as inputs into phenotypic traits at all levels of biological organization. The effects of different inputs interact, going in the same or opposite directions, and they can have additive or non-additive effects on the phenotype of interest” (Jablonka and Lamb 2020, 21).

Yet even if we leave aside the question whether and how these different inheritance systems interact, the obvious point I want to make is that Lamarckism *does* apply to cultural evolution, because its means of replication and transmission differ fundamentally from those of natural evolution. Poverty and lack of education are just as inheritable as are blue eyes and dark skin. The difference is that the latter heritage operates at the molecular-genetic level, whereas the former operates at the socio-cultural level. Both modes of inheritance are effective in their respective domains, though the speed of cultural evolution far outpaces that of natural evolution. Thirty years means nothing in terms of genetic change, yet it means the world in terms of cultural change—the internet, for example, did not exist 30 years ago. And while there is no evidence that digital culture has changed our genetic code (yet), there is plenty of evidence that digitality causes neurophysiological changes during ontogenesis (Carr 2020). Over time, these changes can result in natural selection for greater brain plasticity at the genetic level, as M.J. West-Eberhard (2003) has argued. This leads us to a second observation.

4.2. Cultural evolution is as primary and foundational as natural evolution

The pervasive belief that human culture is somehow a secondary phenomenon that "follows after" biological evolution has been a major obstacle in theorizing cultural evolution. In previous work, I traced this belief back to 19th century philosophy and social theory, like Friedrich Engels' famous claim that "Marx's inquiry begins precisely where Darwin's inquiry ends" (Strathausen 2017, 67). Engel's formulation implies that organic evolution (biology) and human history (culture) are mutually exclusive or separate phases of development when, in fact, both coemerge simultaneously and interact with one another throughout human history. "Human culture participates in *ultimate* causation," Boyd and Richerson rightly insist, because "the evolution of culture has led to *fundamental* changes in the way that our species responds to natural selection" (Boyd and Richerson 2005, 21). In neurological terms, this means that modern cultural evolution "was accompanied by an increase in the plasticity of the brain's neuronal network, which led in turn to a heightened capacity for learning." Since then, Changeux concludes, "cultural evolution [has] taken over from biological evolution" (Changeux 2009, 147)." Human nature is always already culturally coded, because our biological evolution is inseparable from our cultural evolution. At the same time, cultural evolution nonetheless differs from natural evolution. What is that difference? This leads us to a third observation.

4.3. There are no core units of culture[2]

In a series of books and essays co-written with Charles Lumsden during the 1980s, the renowned neo-Darwinian biologist E.O. Wilson introduced the term "culturgen" to develop a theory of cultural evolution distinct from the better-known meme theory (discussed below). "Culturgens," Lumsden and Wilson wrote, "are the basic units of culture", and they are defined as "a relatively homogeneous group of mental constructions or their products. In our classification the manufacture or use of a particular artifact is therefore a culturgen" (Lumsden and Wilson 1983, 121)

This seemingly simple, straightforward definition carries a host of conceptual problems hidden within. First, culturgens are defined as three in-one: (material) product, (mental) process, and (phenomenal) human behavior. Second, Lumsden and Wilson use the term "culturgen" to designate both a "*group* of mental constructions or their products" as well as individual elements *within* that group, such as "the manufacture or use of *a particular* artifact" (ibid; my emphasis). The basic unit of culture is thus defined as both a set and/or an element within a set. This conceptual ambiguity is due to the heterogeneous nature of human culture, which comprises not just human behavior and artifacts of all kinds (rituals and ceremonies, laughter, murder, language, art, tools, tool-making, tool-use…), but also social institutions (buildings, offices, archives, documents, rules, technologies…), as well as mental concepts (ideas, lies, love, hate, dreams, symbols, myths, religions…).

To bring order into this chaos, Lumsden and Wilson argued that some traits of human culture (i.e., culturgens) are common to all cultures and thus generic to the whole species, while others are not. To know which is which, they created monothetic and polythetic sets of entities so as to quantify the degrees of relationship that exist between these entities. A monothetic set of entities is one wherein all entities share the same diagnostic attributes (size, shape, color, duration of a process…), which is the most simple case. An example of a polythetic set might

---

[2] There is no "smallest unit" or "core" of *nature*, either. Reality emerges from dynamically interacting waves and patterns, not static objects. Genes are not entirely autonomous, because their instauration into reality involves distinct patterns and networks in their environment that are irreducible to one another.

be “an array of swords or marriage ceremonies, in which each entity possesses a large number of the attributes of the group” yet “no single attribute is both sufficient and necessary for group membership” (Lumsden and Wilson 1981, 27). Using this method, a culturgen can thus be defined as a monthetic set, or as any polythetic set of cultural artifacts, behaviors, and mentifacts that collectively share a sufficiently high number of diagnostic attributes (at least 45 %) amongst them to be “relatively homogeneous” (Lumsden and Wilson 1981, 29).

The logic behind this model helps to clarify why Lumsden and Wilson had little choice but to define a culturgen as both a set of elements and as elements within a set. For regardless which taxonomy or diagnostic attributes they choose to delineate a particular set (i.e., culturgen), its boundaries always remain open to both higher and lower taxa, because its constitutive elements (i.e., behaviors, artifacts, and mentifacts) are not self-same, essential units. Rather, they are sets, too, determined by clusters of other shared attributes, and so on. A culturgen, in short, is a set within sets of others sets. It comes into being solely as the result of statistical procedures used by Lumsden and Wilson to sort cultural traits into more or less arbitrary groups (>45% homogeneity among its elements) that can be treated as single units and then subjected to further statistical analyses. Yet these analyses have little to tell us about the individual phenotypes they reckon with. On the basis of Lumsden and Wilson’s model, it is impossible to determine whether a particular human behavior, artifact, or mentifact actually belongs to this or that particular culturgen, because culturgens have no objective core or essence, and because the laws of statistics allow me to construct any number of reasonable configurations of diagnostic attributes among and within these diverse (sets of) entities.

The same is true of memes. Recall that Richard Dawkins, in an aside remark in 1976, introduced the term “meme” as the cultural analogion to “gene.” Like genes, Dawkins explained, memes are distinct cultural replicators that give rise to a new and different kind of evolution analogous to, but nonetheless distinct from, genetic evolution. A meme is “a unit of cultural transmission, or a unit of *imitation.* Examples of memes are tunes, ideas, catch-phrases, clothes fashion, ways of making pots or of building arches” (Dawkins 1976).

Dawkins’ formulation was originally met with broad enthusiasm across the field of evolutionary studies. Yet decades of scholarship since then have mainly served to clarify why the gene/meme analogy does *not* hold and how cultural evolution fundamentally *differs* from natural evolution. A major problem was the impossibility to define a “meme” objectively, because a meme can be understood as either

12.1. abstract ideas or knowledge, or

12.2. the nodes or neuronal maps correlated with these abstract ideas or knowledge inside the human brain, or

12.3. the material-symbolic record or representation of such ideas stored on a variety of artificial media or storage devices *other than the human brain* (such as stone, paper, computers), or

12.4. the human behavior or other observable phenomena that either emerge from, represent, or otherwise refer to or express these abstract ideas in human society.

Conceptual equivocations of this magnitude—whereby a meme can be understood as anything from a neuronal net to the Eiffel Tower or the act of lying to Congress--signal that the concept is far too broad and needs to be scaled to size before it can be made useful for a specific discipline or scientific project. The Extended Evolutionary Synthesis (EES) in particular helps clarify that what Dawkins' called a "meme" actually involves a vast set of different modes of transmission and inheritance systems. Many of these systems, we have seen above, do not operate on the basis of strict replication, but aim for less restrictive modes of repetition that are marked by difference and innovation vis-à-vis the original. It is unreasonable and inaccurate to reduce these distinct inheritance systems and processes to a single mode (replication) that supposedly operates on a single kind of being (meme) across all cultures of evolution.

There is no functional equivalent to genes in culture. The common claim that memes are "ideas" or "knowledge-replicators" strikes me as a copout back to Plato. For whatever they are, "memes" must always already have materialized in one medium or another, and the ideal knowledge memes allegedly carry is inevitably affected by these media and their distinct memetic articulations. Everything else is incompatible with the basic premise of scientific materialism.

The slippery ontological nature of memes aside, there were other theoretical problems that meme-theory was unable to solve.[3] Though cultural variants certainly compete with one

---

[3] This includes the fundamental question of whether cultural evolution was adaptive or maladaptive from the gene's point of view, or how to account for the fact that the replication and transmission of memes usually involves a high amount of significant transformations—because I sing a popular tune differently than others, or because I might add something or leave something out—whereas the transmission of genes usually does not (or else replication fails).

another, their mode of competition is less prescribed, literally, than that of genetic variations. It is normal for cultural artefacts to fall apart into smaller units or morph into something else entirely during their transmission or replication processes. Which is to say that popular cultural artifacts will unleash multitudes of different artefacts that vary across a large number of discernable traits (e.g. languages, media, cultures). Given the massive addressability of cultural artefacts, there is no predetermined structural limit to the kinds of variations or adaptations produced during reproduction.

4.4. Literary texts are multi-dimensional objects of study

Let us take a look at Mary Shelley's seminal novel *Frankenstein* (1818) to illustrate this problem. To make things easier, we shall disregard the complicated pre-publication history of this text, including the significant emendations that Percy Shelley, Mary's husband, had introduced into her original manuscript *prior* to its eventual publication—emendations that fuel the ongoing scholarly debate about which parts of the text were actually authored by Mary and which were written by Percy. The first print-edition of *Frankenstein* was released in 1818 in three volumes by Lackington, Huges, Harding, Mavor, & Jones in London and has been disseminated since then in countless reprints and different editions throughout the world. If scholars study this text, or if it gets adapted into a different medium, does it matter which edition is being used? One would think so. Unlike the original 1818 edition, the third edition published in 1831 by Colburn and Bentley was not only an illustrated edition—the first in a long series of illustrated editions to follow—but also featured a preface by Shelley herself, who had carefully worked over and revised the text such that it differed significantly from the first 1818 edition (cf. Mellor 2011). And exactly 100 years after the third edition, in 1931, Grosset and Dunlap in New York City published the first American "movie tie-in" edition that featured a picture of Boris Karloff as Frankenstein (from the 1931 movie *Frankenstein* directed by James Whale) on its front cover. That edition, like many others both before and after, also dropped the original subtitle of the novel. The full title in 1818 had been: "Frankenstein; or, The Modern Prometheus."

Do these subsequent editions of *Frankenstein* still contain the same text? To what degree can the original text of the novel be modified without changing it? Should editors be allowed to correct mistakes or add commentaries, for example? Is a reprint of the same text on different paper with a different font or layout a substantial change of the original text? The answer to these questions depends on the specific objective of the person who raises the question. A rare book collector, obviously, will need to study the original 1818 edition to determine the authenticity of a particular copy she wants to purchase, whereas a media theorist or art historian will likely be content with a facsimile copy of that edition provided it displays all details and other marginalia of the original text. Most students and general readers, finally, will be happy to use just about any available copy as long as the "actual text" essentially remains the same. But what exactly is the "actual text"? Here is how the well-respected editor J. Paul Hunter characterizes the text he chose to publish in the Norton Critical Edition of Mary Shelley's *Frankenstein* (2011):

> The text of Frankenstein printed here is that of the 1818 first edition…. *Only* glaring typographical errors have been corrected; *otherwise* the text reproduced here is that read by Frankenstein's first readers, *except* that explanatory notes have been provided with the needs of modern students in mind. (Hunter 2011, xvii; italics mine).

The editor's palpable hesitations raise the very question he claims to answer: is this the "original" text or not? What is "essential" to Shelley's novel? These questions cannot possibly be answered outside a specific disciplinary framework that determines which aspects of Shelley's text are relevant to which discipline and for what reason. Different disciplines construct different objects of study; that precisely is their scientific purpose and objective. Hence it is impossible to identify any "original" text as *the* source for all subsequent *Frankenstein* adaptations, because different adaptations are based on different versions of the text. There are no originals in the history of cultural adaptation, and there is no way to objectify similarities and differences between a source and its adaptation outside the context of disciplinary paradigms and practices.

"Adaptation is repetition," Linda Hutcheon rightly pointed out, "but repetition *without replication*" (Hutcheon 2012, 7; my emphasis).[4] The crucial point is that cultural evolution operates less by vertical lineage over time than by horizontal spread and diffusion in space. Thanks to digital culture and the internet, there are endless possibilities today for anybody to remix and explore recombinant appropriations of songs or videos on the internet (Voigt 2017). This leads to a final observation.

4.5. Cultural transmission is biased

The evolution of cultural artefacts involves human intentionality, because the cultural transmission of artefacts and their adaptations travel along the choices people make and the concerns they have. There are no cores of culture, because the human mind is an extremely resourceful tool for the creation of ever more, and ever more different, forms of "culture." In the evolution of culture, mentalism matters.

In his comprehensive review of current theories and models for the study of cultural evolution, Wimsatt makes a similar argument: "Cultures are complex beasts, in many ways more analogous to evolving eco-systems, in part of the richness and diversity of modes of horizontal transmission" (Wimsatt 2019, 2). Wimsatt's study confirms that cultural evolution operates primarily through adaptation, no replication, and that it primarily uses flat, horizontal networks as opposed to deep, vertical systems. His multi-partite model of cultural evolution echoes the different levels and systems of inheritance described in the Extended Evolutionary Synthesis (EES). Evolution emerges from the multi-level interactions between these complex systems; it does not emerge from a one-direction flow of information from genotype to phenotype.

II

---

[4] Hutcheon observes that distinction throughout her book, for example by referring to adaptation as "repetition with modification" and not "replication with modification" (Hutcheon 2012, 175).

Let us now switch over to the other side of the two cultures, the humanities. As its name suggests, Adaptation Studies is a genuinely interdisciplinary field in the humanities that examines how cultural artefacts migrate across different milieus or can be refitted into new artistic forms for different audiences. Originally dedicated almost exclusively to film adaptations of novels (Bluestone 1953), the field has grown enormously since then. It now comprises not only literal translations of texts into foreign languages, but also editions and adaptations of texts across diverse formats, genres, media, and cultures.

4.6. "Adaptation" changes its meaning in cultural evolution

The interdisciplinary field of Adaptation Studies has been struggling for decades to define what, precisely, "adaptation" means. A major obstacle has been that adaptation, in evolutionary science, describes both a process and a product. Living beings are the result of previous adaptations that date back to pre-historical times, while at the same time, they also keep adapting to their new current environments. In evolutionary biology, "adaptation" refers to both processes uniformly: an ongoing process ("adapting") and a fixed product of that process ("adapted").

In Adaptation Studies, by contrast, "adaptation" commonly refers to the product of adaptation, not the process, and the object of study is almost always a phenotype, not a genotype. James Whale's cinematic adaptation of *Frankenstein* (1931) is not an adaptation of the medium "text" or the genotype "novel" or the newly emerging genre of "horror." It is, rather, a deliberate adaptation of a particular text and its specific traits into a different medium and artform. Hence most adaptation studies to date consist of a comparative analysis between different media adaptations of the same source. Cultural reproduction, in other words, is based not on replication, but on repetition and remediation (Bolter and Grusin 2000). Many adaptation scholars define "adaptation" precisely in terms of this media transfer as the remaking of "known texts in other art forms" (Andrew 1984, 97).

Here are two recent definitions of adaptation by seminal scholars widely recognized in the field today: Linda Hutcheon defines "adaptation proper" as "an extended, deliberate, announced revisitation of a particular work of art" (2006, 170), while Sarah Cardwell defines adaptation as "the purposeful refitting of material from one artistic context to another" (2018, 13). Both definitions of adaptation emphasize the key traits that distinguish cultural adaptation from biological adaptations. First, they emphasize the importance of the individual/phenotype as opposed to the genotype ("a particular work or art"; "material [in] context"); second, they emphasize the role on human intentionality ("deliberate"; "purposeful"); and third, they replace "replication" with a different kind of process they call a "revisitation" (Hutcheon) or a "refitting" (Cardwell) of an artefact into new contexts.

This ongoing disciplinary struggle to distance Adaptation Studies from evolutionary theory can be traced back to the 1970s, in particular the intense controversy that followed the publication of E.O. Wilson's book *Sociobiology* in 1975. It reflects the deep-seated reservations many humanists harbor vis-à-vis empirical science in general and neo-Darwinian theory in particular. At its best, they consider evolutionary theory synonymous with scientific reductionism and the neo-Darwinian focus on genetic codes and cultural universalism; at its worst, they hold evolutionary theory responsible for the rise of social Darwinism and racism in the 20th century and beyond.

Even more unfortunate was that the few adaptation scholars who did engage evolutionary theory over the last two decades immediately took a wrong (neo-Darwinian) turn and went down the rabbit hole of meme-gene reductionism. In a co-written article in 2007, for example, Linda Hutcheon sought to adapt Dawkins' meme theory to Adaptation Studies simply by replacing memetic "ideas" with memetic "stories" as the "fundamental unit of cultural transmission" (Bortolotti and Hutcheon, 2007, 447). That is a deeply flawed premise, because it leaves the central dogma of MS is place and concedes far too much ground to the false homology between genes and memes.[5]

---

[5] "By homology," Bortolotti and Hutcheon write, "we mean a similarity in structure that is indicative of a common origin: that is, both kinds of adaptation are understandable as processes of replication. Stories, in a manner parallel to genes, replicate" (Bortolotti/Hutcheon, 2007, 444). No, they really don't. As I argued in Part I, the repetition process that governs cultural evolution is fundamentally different from the replication process that governs biological evolution. Yet to explain this difference would require a more detailed account of the limits of meme-theory than Bortolotti and Hutcheon provided in their essay. Instead, the authors oscillate back and forth between allegedly synonymous definitions of adaptation (defined as some kind of "memetic replication") that are anything but. Take this claim for example: "We are not saying that cultural adaptation is biological; our claim is more modest. It is simply that both organisms and stories 'evolve'" (446). I agree. But this is fundamentally different from saying that stories "replicate parallel to genes" and share "a common origin" with them. These last claims do not leave much room to explore the distinctiveness of cultural evolution apart from natural evolution.

Hutcheon's intervention did not leave a lasting impression in the field. That is not surprising, given that Adaptation Studies keeps producing massive empirical evidence that testifies *against* the existence of singular core units as the sole agent of cultural evolution. Hutcheon herself, in fact, had pointed that out earlier when she declared that "adaptation is repetition, but repetition without replication" (Hutcheon 1992, 7). Instead, Adaptation scholars recognize these so-called core-units of culture for what they are: relatively arbitrary historical crystalizations of multi-dimensional forms and figures that gradually change over time.[6]

4.7. Cultural Adaptation exceeds Intertextuality

Rather than building on Hutcheon's original insight and keep exploring the difference between cultural repetition and biological replication, however, the field of Adaptation Studies over the last 50 years has moved in the opposite direction, basically ignoring not just neo-Darwinian dogma, but evolutionary science and systems theory altogether. In its lieu, scholars keep suggesting different taxonomies and classification systems that no longer make much reference to "adaptation" at all. In 2006, for example, Julie Sanders proposed the following list of terms to describe the different kinds of cultural transformations studied by adaptation scholars: "a version, a variation, an interpretation, a continuation, a transformation, an imitation, a pastiche, a parody, a forgery, a travesty, a revaluation, a revision, and a rewriting" (Sanders, 2006, 18). While this list may seem unwieldy, it pales in comparison to a more comprehensive list of suggestions made by various scholars across the field over the last two decades:

> borrowing, stealing, inheriting, assimilating, influence, inspiration, dependency, indebtedness, haunting, possession, homage, mimicry, travesty, echo, allusion, and intertextuality, ..., variation, version, interpretation, imitation, proximation, supplement, increment, improvisation, prequel, sequel, continuation, addition, paratext, hypertext, palimpsest, graft, rewriting, reworking, refashioning,

---

[6] Guy Spielman introduced the term "crystalization" in his paper on March 13, 2021 at the NeMLA conference.

> revisioning, re-evaluation, bricolage, and pastiche, ..., reading, rewriting, critique, translation, transmutation, metamorphosis, recreation, transvocalization, resuscitation, transfiguration, actualization, transmodalization, signifying, performance, dialogization, cannibalization, re-envisioning, incarnation, and re-accentuation. (Elliott 2020, 182).

The conceptual heterogeneity of this list is mind-boggling… at least from a disciplinary perspective. On the one hand, the list testifies to the rich conceptual heterogeneity at work in Adaptation Studies, and it hints at the multiplicity of different perspectives and interrogations necessary to make sense of cultural adaptations that move across time and space. Each term on this list, after all, has emerged from *empirical* case studies in the field that tried to conceptualize *existing* adaptation practices as best they could. On the other hand, the list is also symptomatic of an academic field that—for better or worse, by accident or by principle—lacks an overall unifying method or set of principles at its core. There is little that holds Adaptation Studies together in theoretical terms, and the field to this day consists largely of isolated, individual case-studies that do not rise to the level of exemplarity necessary to develop an overall theoretical framework or core premise.

At best, Kamilla Elliott points out, adaptation scholars across the field share a vague commitment to the theory and practice of "intertextuality," yet without clear agreement on what, exactly, intertextuality means in methodological or practical terms (Elliott 2020, 165f.). And though some scholars use the term "intertextuality" as synonymous with "adaptation," others worry that an unrestrained notion of intertextuality will lead to a bad infinity of texts that undermines the text/context distinction and renders the study of adaptation impossible or meaningless (Leitch 2012)—or requires computer modeling, as I shall suggest below.

In any case, I agree with Leitch that distinguishing text from context remains a conundrum in Adaptation Studies. But I disagree with him and others who say it cannot be done at all or serves no purpose whatsoever. True, we cannot distinguish between text and context in a *universalist, absolute, and objective manner*, because texts (like artworks) are infinitely addressable, as we noted earlier. In literary studies, the scale ranges from single words and phrases on the page to more abstract traits such as literary style, form, and genre all the way to the large-scale abstractions we call realism, avantgarde, (post-)modernism, and Literature (with a capital L). This wide range of concepts cannot be squeezed into a single category called "text," nor should we try to reduce its different milieus to a single environment called "context." But scholars can, and should, use distinct parameters to define texts provisionally for a particular purpose and within a specific context and disciplinary framework.

Steeped in 20th century semiotics and linguistic theory (i.e., Todorov, Bakhtin, Kristeva), "intertextuality" is too antiquated and one-dimensional of a concept to account for the dynamic networks and remediation processes that define adaptation practices in the digital age. The key point is that not everything outside a given "text" is part of its immediate "context," because text and context are scalable entities that must be clearly delineated if they are to resonate productively with one another. To become productive, the text/context relation must be unfolded and conceptualized differently across the fields, while, at the same time, these differences must recalibrate one another during the course of the inquiry. The resulting networks, of course, will be neither universal nor objective. They will local, concrete, and provisional—yet all the more revealing and productive because of that, in my view.

4.8. Adaptation Studies must adapt evolutionary theory

Like other adaptation theorists, Kamilla Elliott bemoans the profound "dissonance between adaptation and theorization" that characterizes her field. The dissonance, she claims, cannot "be resolved by changing old theories for new," but requires "a revolution in what theorization is and does in the humanities and in its relationship to adaptation" (144). Above all, Elliott concludes, "adaptation studies needs its own principles, based in its own adaptive processes" (230).

I fully agree that Adaptation Studies needs a more adaptive approach to theorizing cultural adaptation. The overall goal should be to "make Adaptation Studies adaptive," as Brian Boyd put it (Boyd 2017). But that requires above all to leave behind the legacy of sociobiology and neo-Darwinian reductionism. Instead, a good taxonomy for the study of cultural adaptation should recognize these mythical core-units of culture for what they are: relatively arbitrary historical crystalizations (Guy Spielman) of multi-dimensional forms and figures that gradually change over time within the dynamic network of fiction.

The best way to make Adaptation Studies adaptive is for scholars to engage constructively with contemporary evolutionary theory instead of ignoring it. The Extended Evolutionary Synthesis, I have argued, offers a rich conceptual and methodological toolbox for the study of cultural adaptation that has nothing in common with neo-Darwinian reductionism. EES offers far more to the humanities than just-so-stories about replicating memes and selfish genes. The system-biologist Terrence Deacon, for example, is a strong proponent of biosemiotics, the idea that the bio-chemical processes that define life always already include, or can be conceived as, semiotic processes. Life is at one and the same time both chemical and semiotic in nature, Deacon and his colleagues explain in a joint manifesto:

> An aim of the biosemiotics approach is to explain how life evolves through all varieties of forms of communication and signification (including cellular adaptive behavior, animal communication, and human intellect) and to provide tools for grounding sign theories. We introduce the concept of semiotic threshold zone and analyze the concepts of semiosis, function, umwelt, and the like as the basic concepts for theoretical biology. (Kull et al 2009, 167)

It should be obvious that biosemiotics holds far more promise for theorizing Adaptation Studies than genetic reductionism. A key concern of biosemiotics is to push back against the neo-Darwinian mythologization of genes as autonomous agents for the creation of life. Instead, Deacon's research documents the diverse biological environments and epigenetic constraints that contribute to gene replication, gene function and gene expression at the molecular level. Like Jablonka and Lamb, Deacon supports the extended evolutionary synthesis and recognizes the relative independence of different inheritance systems. He emphasizes again and again the importance of intra-evolutionary mechanisms at the ontogenetic level that operate in conjunction with evolutionary mechanisms at the phylogenetic level.[7]

If we transfer Deacon's insights from the biological to the cultural level, we might say that adaptations of cultural artefacts traverse multiple levels of distinct environments with variable selection criteria and selection pressures at each level. Some of these adaptation networks may aim for fidelity to the source, while others may prefer dissimilarity to increase their own visibility. To distinguish them, Adaptation Studies needs to adapt some key evolutionary concepts—concepts like semiosis, system, function, umwelt--into the disciplinary milieus of

---

[7] "The replication, variation, and differential preservation that together characterize natural selection have their counterparts in the redundancy, degeneracy, and functional interdependencies that characterize intraorganismic processes" (Deacon 2010, 9003).

the humanities. The resulting definitions and theoretical frameworks should aim to provide basic heuristic tools for critical analyses of cultural adaptation processes—no more, no less. They should serve as better mouse traps that allow scholars better to distinguish relations and networks within the rich and complex history of adaptation practices that operate across different languages, cultures, media, and disciplines. However popular it remains in Adaptation Studies today, intertextuality is no longer an adequate term to describe or analyze these processes.

4.9. Summary

Neo-Darwinian cultural theory failed for two reasons: first, because its basic assumption about cultural evolution is deeply flawed. Cultural evolution cannot be reduced to essential core units whose replication process is analogous, or homologous, or functionally equivalent to how genes replicate in natural evolution. There are as many core-units of culture as there are people concerned about them. Any-thing can be conceived of or function as a replicator in human culture. The links and networks that tie cultural adaptations to their many sources are infinitely more varied and volatile—because they are far less rule-bound—than the bio-chemical networks prescribed by genetic and epigenetic rules for living systems. Finally, cultural evolution runs along human intentions, whereas natural evolution does not.

The second reason why the neo-Darwinian model failed had less to do with its theory and more with the lack of concern and diplomatic grace among its proponents. As I argued in detail elsewhere (Strathausen 2017), a strictly scientific-reductionist approach to culture is bound to overlook, or forced to ignore, a good-many objects of concern that are widely cherished across the humanities *precisely because* they are irreducible to quantitative methods of inquiry. These objects include the haecceity (the singular "here and now") of the artwork, the phenomenology of aesthetic experience, and the method of close reading, among others.

The study of cultural evolution requires a broad, interdisciplinary approach that spans the gamut from quantitative computer-modeling and statistical approaches on one end to qualitative and speculative approaches based on narrative on the other (Love and Wimsatt 2019). The two extremes at the end of the spectrum have proven unproductive and serve mostly to delineate the methodological parameters across the field. Neither a purely statistical, nor a purely narrative approach will achieve meaningful results; only a combination of the two carries any promise at all for the study of cultural evolution.

References

Andrew, D (1984) Adaptation. In Dudley Andrew (ed) *Concepts in Film Theory*. Oxford UP, 96– 106.

Bluestone G (1957) *Novels Into Film*. Johns Hopkins P, Baltimore.

Bolter DJ, Grusin R (2000) *Remediation: Understanding New Media*. Rev. ed. MIT UP, Boston.

Bortolotti G R, Hutcheo L (2007) On the Origin of Adaptations: Rethinking Fidelity Discourse and "Success"—Biologically. *New Literary History* 38(2007):443–58.

Boyd B (2017) Making Adaptation Studies Adaptive. In Leitch T (ed) *The Oxford Handbook of Adaptation Studies*. Oxford UP, New York, pp. 637-57.

Boyd R, Richerson P J (2005) *The Origin and Evolution of Cultures*. Oxford UP, Oxford.

Cardwell S (2018) *The Routledge Companion to Adaptation*. Routledge, London and New York.

Carr N (2020) *The Shallows. What the Internet is Doing to Our Brains*, 2nd ed. Norton, New York.

Changeux JP (2009) *The Physiology of Truth. Neuroscience and Human Knowledge*. Harvard UP, Cambridge.

Deacon T W (2010) A role for relaxed selection in the evolution of the language capacity. *PNAS* 107/2(May 2010):9000-6.

Elliott K (2020) *Theorizing Adaptation*. Oxford UP, Oxford UP.

Hayles N K, Pressman J (eds) (2013) *Comparative Textual Media*. U of Minnesota P, Minneapolis.

Hunter J P (2011) Introduction. In J. Paul Hunter (ed), *Mary Shelley. Frankenstein*. Norton Critical Edition. 2 ed. Norton, New York pp. ix-xviii.

Hutcheon L (2012) *A Theory of Adaptation*, 2nd ed. Routledge, London & New York.

Jablonka Eva, Lamb Marion (2020) *Inheritance Systems and the Extended Synthesis*. Cambridge University Press Cambridge.

Kull K, Deacon T W, Emmeche C, Hoffmeyer J, Stjernfelt F (2009) Theses on biosemiotics: Prolegomena to a theoretical biology. *Biological Theory* 4/2(2009):167-73.

Leitch T (2012) Adaptation and Intertextuality, or What Isn't an Adaptation, and Does It Matter? In Cartmell D (ed) *A Companion to Literature, Film, and Adaptation*. Wiley-Blackwell, Oxford, pp. 85–104.

— (2017) Against Conclusions. In Leitch T (ed) *The Oxford Handbook of Adaptation Studies*. Oxford UP, New York, pp. 755-67.

Love A C, Wimsatt W. C. (eds) (2019) *Beyond the Meme. Development and Structure in Cultural Evolution*. U of Minnesota P, Minneapolis.

Lumsden C, Wilson EO (1981) *Genes, Mind, and Culture*. Harvard UP Cambridge.

— (1983) Promethean Fire. Harvard UP, Cambridge.

Mellor A K (2011) Choosing a text of Frankenstein to teach. In J. Paul Hunter (ed), *Mary Shelley. Frankenstein*. Norton Critical Edition. 2 ed. Norton, New York, pp. 204-11.

Strathausen C (2017) *Bioaesthetics. Making Sense of Life in Science and the Arts*. U of Minnesota P, Minneapolis.

Voigts E. (2017) Memes and Recombinant Appropriation: Remix, Mashup, Parody. In Thomas Leitch (ed), *The Oxford Handbook of Adaptation*. Oxford UP, Oxford, pp. 308-27.

West-Eberhard M. J. (2003) *Developmental Plasticity and Evolution*. Oxford University Press, New York.

Wilson E.O. (1975) *Sociobiology. The New Synthesis*. Harvard UP, Harvard.

Wimsatt W.C. (2019) Articulating Babel. A Conceptual Geography for Cultural Evolution. In Love and Wimsatt (eds) *Beyond the Meme. Development and Structure in Cultural Evolution*. U of Minnesota P, Minneapolis, pp. 1-41.

Chapter 5

The epistemological and ideological stakes of literary Darwinism

Alexandre Gefen[i]

Abstract

Proposing to produce “new humanities,” the literary Darwinists affirm the failure of the humanities in the face of modern scientific demands. Literary Darwinism promises to naturalize literary aesthetic practices, both poetry and narrative. It justifies the human need to produce fiction and proposes to interpret its contents. By confronting cultural studies head-on, as well as previous paradigms, from psychoanalysis to historicism, evolutionary theorists have opposed postures defending the insularity of art and the differential character of artistic practices. Such a position has been exposed to virulent criticisms (reductionism, conservatism, utilitarianism, essentialism, scientism, etc.) to which I would like to return, as they seem to me to hinder a serene examination of the disciplinary proposals put forward, in all their richness and epistemological ambitions.

---

[i] CNRS and Université Paris 3-USPC, Paris.

The cognitive sciences constitute a scientific domain rather than a single science. Originally organized according to a triad constituted by experimental psychology, neuroscience and information theory, they now deploy a variety of levels of analysis and fields of expertise, whether they are interested in the molecular, cellular, neuronal, behavioral, social or even anthropological supports of human cognition. Placed in a very particular position within the sociology of science, since they are characterized both by their internal heterogeneity and by the ubiquity of their questions, the cognitive sciences have been accused of various evils (positivism, dualism, formalism), and, in reaction, have recently developed in the direction of the analysis of complex phenomena and in original orientations that make them much more than a description of cerebral functioning: embodied cognition, affective sciences, social cognition or shared cognition. Starting from the idea that all the human facts, including the facts of imagination, are included in a process of evolution, the evolutionist conception of the aesthetic mechanisms consists in putting aside the metaphysical, sociological, economic, formal, psychological interpretations (in the non-evolutionist sense of the term), to wonder in what the aesthetic representations illustrate, exemplify or model the play of biological forces "*hard-wired*" (Max, 2005): the survival, the reproduction and the expansion of the species, the competition and the cooperation between the men, the families and the communities, the kinship, the social affiliation, the efforts to acquire resources and influence, the domination, the aggression, finally the need of imagination (Carroll, 2011b: 30). To think with evolutionary psychology is also to ask how the production of aesthetic objects — man, says Jonathan Gottschall, a great figure of this emerging discipline, is "a *storytelling* animal" (Gottschall, 2012a) — participates in the nature of man as a species and in his evolution, whether it is a question of explaining the aesthetic aptitudes as a parasitic biological competence, as an elaborated aptitude of adaptation to an environment, or still, as Jean-Marie Schaeffer does, as an analogue of the mechanisms of optimization of the choice of the partner in the sexual reproduction (Schaeffer, 2009).

The evolutionary paradigm offers a theory that is both very simple and powerful in its principles, much less technical and much more powerfully heuristic than literary neuroscience; it is, moreover, and I will come back to this, very heavy with implications, as optimistic as they are troubling, as to the place of culture and artistic practice in our societies, and as to that of the humanities in the map of knowledge and the academy: conceived as an indispensable natural adaptive capacity, literature becomes inseparable from man's humanity, it acquires a function and therefore an indisputable legitimacy. If, after a 19th century marked by history, a 20th century marked by the triumph of linguistic knowledge, it is to a cognitivist 21th century that we will be confronted, and it is in my opinion first of all with evolutionary psychology that literary criticism and theory will have to dialogue.

5.1. Epistemological situation of evolutionary psychology

Before returning to these theses, I would like to make a few general remarks of an epistemological nature, without which it seems difficult to understand the place of these theories and which have to do with the epistemological status of the cognitive sciences and of evolutionary psychology in general. Cognitive evolutionism is in fact part of both philosophical naturalism and scientific positivism: the ultimate substratum of mental facts is physical, natural, and the ultimate determinisms are genetic. The theoretical anchorage of understanding of our knowledge, its foundation, must be that of human cognition, as a natural phenomenon explicable *in fine* by the sciences, without recourse to external metaphysical or linguistic models: we are in the framework of what Quine calls a "naturalized epistemology", to which it would be permitted to employ "the resources of natural science" (Quine, 1969b: 90). In this total Spinozism, which refers to a form of critical materialism (which, in a sense, is the equivalent for our generation of post-war Marxism) (Guerin *et al.*, 2010: 146), there is only a strictly biological substratum to mental facts, and cerebral evolutions are determined by mechanisms proper to the augmented evolution of Mendelian theory: random individual variation, environmental pressure, genetic selection and transmission of optimal adaptive characteristics. Here, consciousness, religion as well as literature are products of neuronal evolution governed by the need to master a complex and dynamic human environment — unlike instinctual behaviors or a simple adaptation to a fixed and repeated environment, attributed to *Homo sapiens*: the characteristic of man as a species is cognitive fluidity and neuronal plasticity (Mithen, 1996).

The consequences of such a movement of "naturalization of the human being" (Lageira, 2012: 50) are considerable and it is important to specify them. First of all, we should note that this naturalization of aesthetics is not comparable to a return of "psychologism" in the traditional sense, or to the theory of an evolution of literature in the sense of Brunetière (who, as we remember, applied the Darwinian model to the logic of literary genres [Brunetière, 1890;

Compagnon]). As with Quine, whose naturalization of epistemology constitutes a major framework of analysis, we are in the framework of a psychology considered "in a non-psychological way" (Laugier, 2010: 24), that is to say an anti-intentionalist, anti-mentalistic psychology, where there is no need to "interpret" private expressions in a supposed abstract "mental language" that would have its logic, but simply to classify empirical facts according to a deductive scientific logic. If we take again an opposition posed by Wilhelm Dilthey between natural sciences *(Naturwissenschaften)*, producing explanations, even predictions, with the help of experimental demonstrations or mathematizable reasonings, and human sciences *(Geisteswissenschaften)*, producers of comprehension, the evolutionary psychology of the literature has vocation to be reintegrated in sciences in general and to produce not *interpretations*, but *explanations* of aesthetic facts. It is a question, when one speaks about the psychology of the art, to found it on the empirical observation, even on the experimentation, and to guard against any endogenous speech. Some hard theorists like Harold Fromm (Fromm, 2009) are in fact close to what is called "*eliminative* materialism," represented in particular by Daniel Dennett[1] — who refers the mental facts to a physiological substratum beyond the reach of common sense, including the *qualia* (subjective effects of our sensations and experiences), and Richard Dawkins[2] — the inventor of the meme theory, that is to say of a conception of the artistic facts and of the culture in general as "foldable units, supports of formal elements or stylized contents" (Morizot, 2012: 185), subjected to the laws of the natural selection. Without going so far, the "sociobiological" thought to take back a concept very used by the "literary Darwinists"[3] tends to refute the interpretation (at least the internalizing interpretation), the speculation on the particular and individual value of the works or on the personality of the author, in the name of the logic of the very long duration (the starting point of the cognitivist literary history, it is not Homer, but the "'Human Revolution,' somewhere between 100,000

---

[1] Dennett is professor of philosophy at Tufts University, where he co-directs the Center for Cognitive Studies. On his theory of mental properties, see Dennett, 1991;on his adaptationist views, Dennett, 1995; on his atheism, Dennett, 2006.

[2] Dawkins is a biologist and ethologist and Professor Emeritus at New College, Oxford. See Dawkins, 1976; Dawkins, 1982; Dawkins, 2006.

[3] Expression coined by the American biologist Edward O. Wilson. See Max, 2005.

and 30,000 years ago" [Carroll, 2011b: 26]) or of the quantitative logic. For Jonathan Gottschall, "one thing literature offers is data. Fast, inexhaustible, cross-cultural and cheap" (quoted in Max, 2005). One can think here of the concept of *big data* in the digital world: the amount of empirically available data is so large that its quantitative examination would supplant any theory. It is a question, Gottschall continues, of producing "literary hypotheses that make testable predictions about empirical reality" (Gottschall, 2008a: 64). It is then a question of proposing an empirical description, if not experimental, of the art, ambition which reverberates to other disciplines and notably, if one believes Gerald Prince, to the post-classical narratology (Prince, 2008) and which leads to forms of description of the literary fact which often tend to refuse any speculation on the particular cases, private, in the name of general and massive facts or, when it is interested in singular facts, to avoid any aesthetic speculation in favor of a simple description of interactions supposedly governed by the necessities of the species and making sense in series. Hence what Carroll considers an epistemological modesty: literary texts cannot be "deciphered" in another code, because evolutionary theory assumes that texts are cultural tools that operate on ordinary behaviors with ordinary language, that they incorporate a "folk understanding of human nature" (Carroll, 2011b: 110, see 29; Guerin *et al.*, 2010: 146). If "most texts are understood reasonably well at the level of common language and common knowledge," they must therefore be interpreted through common language (Carroll, 2011b: 29). Darwinians, unlike poststructuralists, therefore forbid the use of "prefabricated sign systems" in which they would translate the content of texts (29). In the same way that we try to avoid what we could call among the Wittgensteinians the illusion of interiority, we are going to avoid what Quine called "the museum myth" of interpretation (Quine, 1969a: 27 ff.) and the idea that meaning is part of a discourse that goes beyond the data we have to speak about language, by reinscribing on the contrary in natural behavioral determinisms the textual representations, which have neither mental substance, nor mechanics of their own (Laugier, 1992: 93).

Another consequence of this epistemological position is that there would be no more formal logic or specific historicity to symbolic productions than ontological or even functional autonomy of literature, which loses all specificity. There is no more reason to circumscribe an "empire within an empire," and, to say it with another vocabulary, the cognitive evolutionism integrates itself to a general paradigm that Jean-Marie Schaeffer qualifies of "end of human exception." It consists in refusing any anthropocentric, teleological or essentialist conception of man (Schaeffer, 2007: 185-200), in order to examine the so-called "humanity" "in the light of the constraints that govern it as a biological hope" (201). Let us note in passing that for Schaeffer, as for all the theorists of evolutionism, there is no opposition between cultural order and natural order, since culture is a "natural" attribute of the human species; it is on the contrary a question of admitting the existence and of showing the sense of the interactions between biological dispositions and cultural dispositions — what is called the genetic-cultural coevolution (341-343) — or between resources and collective constraints, on the one hand, and particular individual aptitudes, on the other hand (249). It should be noted here that this anti-essentialism is to be distinguished from historicism in the traditional sense of the term as well as from the philosophies of existence: if man has no essence, it is in so far as he is an animal, a biological and genetically modellable being, constituting lineages in constant evolution, and not in so far as he transcends by his freedom his biological identity (198-199). It is thus a question of not seeking the ultimate reasons of the symbolic textual productions in the cultural standards and their determinations, but in a "human nature" of genetic order which includes the culture. One sees here the power of disruption of these theories: it is by the natural adaptation that we must account for the ontology of the representations, it is by it that the link between literature and reality is made. Joseph Carroll, leaning on the ethologist Konrad Lorenz, thus affirms that "the human senses and the human mind have access to reality because they have evolved in adaptive relation to a physical and social environment about which the organism urgently needs to acquire information" (Carroll, 2011b: 20).

5.2. The debate on the adaptive function of literature

The first and perhaps most interesting debate concerns the functions of literature conceived as a response to a need for adaptation. Several theories have been proposed in response to the paradox posed by the emergence of art among properly human activities: how is it that the human species has spent so much time on useless and altruistic activities such as fiction, instead of hunting and reproducing? According to Gottschall (2004b: 12), the will to propose a non-metaphysical explanation to the necessity of art has attracted many eminent evolutionists, who see in the production and consumption of art either an adaptive result of natural selection (Wilson, 1998; Tooby and Cosmides, 2001), or an adaptive result of sexual selection (Miller, 2000) or a non-adaptive by-product (Pinker, 1997; Buss, 1999: 407-410) of this same sexual selection. This debate between three possible explanations of the human artistic fact can be found in most of the typologies proposed by evolutionists, in Carroll (2011b: 20-29, 49-53) as well as in Brian Boyd (2005), who opposes a conception of art as *adaptive* attention, another of art as a *by-product* of evolution and a last one making art a mechanism directly derived from the processes of sexual selection *(sexually selected)*.

5.2.1. Theory of enjoyment, in other words of art as a parasitic activity

Let's start with the theory of art as a by-product, sometimes called the theory of art as an *spandrel.*[4] For Steven Pinker (Pinker, 1997; Pinker, 2007; see Mellmann, 2011: 315-316), art "presses" our pleasure buttons, designed for other purposes. It would be analogous to our need to consume pornography whereas our needs are only reproductive, except that the pressed buttons are not those of the sexual activity. Hence the image of *cheesecake*, a pastry elaborated to titillate the sensitive points of our mental faculties: we like cheesecake because we have developed not a taste for it as such, but circuits that trigger in us a flow of pleasure when we experience the sweetness of a ripe fruit, the creamy sensation in the mouth of fats and oils, the freshness of water. "Cheesecake packs a sensual wallop unlike anything in the natural world because it is a brew of megadoses of agreeable stimuli which we concocted for the express purpose of pressing our pleasure buttons. Pornography is another pleasure technology. At least to some extent, art may be a third" (Pinker, 1997: 524-525; see the review by Carroll, 1998). One sees here the departure made between the necessary capacities, as the sociality and the language, and the optional capacities as the imagination. Art would be thus, like masturbation, the decoupling of a *hardware* faculty of the man in a different *software*, to borrow frequent metaphors among the cognitivists. This position, which can make one smile, has been frequently criticized, by Brian Boyd in particular (2005: 155 ff.), who alleges on the one hand that it totally neglects the public and thus the engagement in art in the form of a social pressure

---

[4] "Gould and Lewontin have proposed a famous analogy between Gothic architecture and evolutionary logic, both of which involve reorientation phenomena; just as the small bones of the inner ear of vertebrates have their origin in the jaw of reptilians although they fulfill a completely different function, the triangular and curvilinear spaces (pendentives or spandrels) between an ogive and a pillar or a dome cannot be avoided as soon as one adopts the gothic style, but the mosaicists of San Marco were able to transform these inconvenient areas into sumptuous decorative supports that were not originally intended," explains Morizot, 2012: 185.

and, on the other hand, that it forgets to what extent art can be an effort). As debatable as it is, this archaeology of the art is implicitly summoned every time we make of a genre or an artistic practice the transposition, the autonomization or the extension of instinctual or natural pre-coded functions. One will find it, for example, in the concept of "exaptation" proposed by Morizot as a possible analogy to understand the way in which the art is born, that is to say the use of an organ for another use that the biologically determined one, theory that he calls indirect and for which "the result," in other words the artistic fact, "is not selected for itself, but as a correlate of a more fundamental reason" (Morizot, 2012: 185).

5.2.2. The theory of expensive signals

The conception of art as a mechanism directly derived from the processes of sexual selection is based on a concept introduced by Darwin in *The Descent of the Species* (1871), "to justify the existence of paradoxical forms of evolution which at first sight seem like handicaps (the tail of the peacock, the antlers of the deer, etc.). The idea is that these disproportionate appendages do not function for direct adaptation (like camouflage), but as a manifestation of potential reproductive superiority" (Morizot, 2012: 185). This idea, enriched by the contributions of Mendelian genetics, can be found in what is called the "theory of costly signals," imported by Jean-Marie Schaeffer in a 1997 article and then in a 2009 book: "The central hypothesis of the theory of costly signals is that the cost or benefit (for the signaler) of this type of disabling signal depends on the *actual* qualities of the sender. The greater these qualities are, the less costly the signal is for him; the smaller they are, the greater the cost." (Schaeffer, 2009: 35) In other words, to quote Schaeffer again, "a costly signal is a signal that cannot be simulated. If one is able to produce it, it is because one actually possesses the qualities that it signals, because it is precisely these qualities that make its production possible" (35). The analogy between costly signals and artistic processes is thus largely due to the fact that "the main stake in the communication of the costly signal is the very existence of this signal beyond anything to which it may otherwise refer" (Schaeffer, 2012: 30), we are here close to what Wittgenstein suggested when he affirmed that the work of art "does not want to transmit something else, but only itself" (Schaeffer, 2009: 47): whether it is a question of sexual selection or of works of art, costly signals have a metasignaletic function. Thus, the sex parade of the cradle bird "functions neither as a simple artifactual construction nor as a simple interaction by hetero-referential signals, but within a self-referential dynamic: the cradle is referred to itself as a decorated structure that materializes the *fitness of* the male, the dance and the calls are self-presentations that both signal and enact its value. This logic is that of the work of art" (47-48). Better, Schaeffer suggests that the contexts of emergence are similar: "What we call *art*, they are the facts of costly signaling, or rather their crystallization in the form of a certain number of typical productions related transculturally (dances, ornaments, sculptures, verbal productions, pictorial representations, etc.) which answer these situations of problematic communication. And what we call *aesthetic relation* is nothing else than a reception itself costly of these signals" (Schaeffer, 2012: 31). The art would be a possible answer to our need of reflexivity in complex or "problematic" existential or social situations, by distant derivation of functionally different adaptive mechanisms. It is undoubtedly that the recourse of Schaeffer to the animal model is worth less as an operational explanatory framework than as a reminder of the unified epistemological order in which it falls to us to think the aesthetic fact.

5.2.3. Theories of adaptive gain

The theories that make of art an adaptive disposition with direct benefits for the human species proceed from a functionalist thought of the aesthetic defunctionalization: they share the idea that the detour of the immediate action (of enjoyment, of sexual reproduction) favors capacities of complex organization of the behaviors and the interrelations. We could, in my opinion, divide them between two big currents, which have in common to have largely reflected on the adaptive role of the fiction. To continue the analysis of Boyd who, in his thought of art as "adaptive" attention (Boyd, 2005: 151 ff.), underlines, on the one hand, a dimension of adaptation and social regulation, and on the other hand, a more general capacity of mental organization (167) offered by the fictional immersion in possible worlds, we could differentiate the theories centered on the social role and the theories centered on the cognitive role of fiction.

5.2.4. Literature as a behavioral adaptation capacity

It is a question of making fiction, in a more or less directly behaviorist perspective, a tool of adaptation and regulation of social behaviors. The capacities of social adaptation induced by art are not then directly linked to the selection of the partner, but to the good functioning of the society: the education to others and more largely the simulation of "a variety of social relationship, behaviors, and consequences" (Sugiyama, 2005: 188) allowed by art allow a better knowledge of the human environment by the exercise of the anticipation and the changes of mental roles. The art and in particular the representative art, allows the species to produce and to transmit a general knowledge on itself, knowledge whose determining specificities are to be concrete, affective, incarnated, and thus to supplement more abstract orders of knowledge. This "knowledge-how", to use Ryle's formula (Gefen, 2010), has not only the strength to produce regulating models of reference and to put on the table normative conflicts that are common in evolved societies (fidelity to the model of desire *vs.* fidelity to the idea of family, obedience to the state *vs.* preservation of the family unit, etc.), but also to explore new ways of thinking and acting.), but also to explore particular cases, even atypical, amoral or traumatic situations on which it informs the group (it is the exemplarity of the inexemplarity [Gefen, 2007]).

One cannot live without a theory of how another's mind works, these philosophers argue, a profound knowledge that literary fiction would stimulate. Not only does literature play a role in the way literature allows us to refigure and appropriate our experience, as narrativist philosophies such as that of Paul Ricoeur have shown, but it would also participate in this fundamental aptitude that is our capacity to envisage the psychological structure of others by acquiring, at least in a summary way, a "cartography of the human mind" *(mind mapping)* and by being able to anticipate the reactions of other individuals (*mind reading*, understanding others). This form of initiation to the otherness is undoubtedly close to what the literary tradition had named according to the formula of Dorrit Cohn "interior transparency" (Cohn, 1978), that is to say the postulation of the literature to discover and to explain the psychic life by the mediation of fictional interiorities. Literature would equip us with cognitive tools of behavioral modelling that allow this form of divination by which we ordinarily interact with the minds of others. This perspective explains, among other things, why literature gives itself as a recurrent theme the psychological portrait or the biographical narrative: such is the framework of many modern plots on which evolutionary thought intends to cast an overhanging glance by likening them to a form of psychological speculation (Guerin et al.; 2010: 149-150; Nettle, 2005). Alan Palmer connects this work of *mind mapping* with the games of enunciative position in the literature which would put in scene the conscience at work in the novel ("in essence, narrative is the description of fictional mental functioning" [Palmer, 2004: 12; see 177]), to the profit of a cognition, an action and even a form of distributed identity (15): for him as for Mark Turner (1997), the human brain considered as an active and plastic system modified by experience and language conceived as an instrument used by separate brains to exert biological influence on others at a distance, to the point of creating a kind of virtual brain of the species that exerts a distributed biological influence. The hypotheses necessarily made by the readers, the conjectures, the reflections that he has to deploy to analyze the situations, the enunciative games of the literature make of the narratives devices to understand the springs of action and cognition of the mind. Lisa Zunshine continues this reflection by adapting it to the analysis of cinema (2008) and, in “Style Brings in Mental States”, refines the “theory of mind” to demonstrate that not only the enunciative effects, but the effects of style itself, participate in this work of understanding the “sociocognitive complexity” of the world of human societies (Zunshine, 2011: 350, 353). *Mutatis mutandis*, one will find an equivalent thought in Carroll: art allows the access to a “total cognitive order” (Carroll, 2005: 87) which is an added value to the world. He produces an analysis of the “life cycle” (83) with the help of literature, allowing the construction of a “cognitive behavioral” framework of reference (92). The notion of collective intelligence seems to me to be particularly interesting: it can be found both in Palmer’s work, who thinks largely in terms of social cognition (Palmer, 2004), and in Terrence Deacon’s work, who affirms that the integration of common histories with the individual histories of individuals allows for the construction of a shared world that is essential to the species (Deacon, 1997).

5.2.5. Art as an added value

Another theoretical option consists in underlining less the cognitive gains than the existential gains for the species and makes of the fiction a mode of resistance to the metaphysical difficulties (the man is an animal conscious of his own mortality) that social (the complexity and the fluidity of the human societies). For Ellen Dissanayake, *Homo Aestheticus* (1992), art is a result as fundamental of the evolution as nutrition, because it gives interest to the ordinary activities of the existence and allows to make the difficulties of the life more livable. Thanks to art and its capacity to instill desire and beauty in the interactions between individuals, the group becomes bearable; thanks to it, human activities are not reduced to finalities such as reproduction and the quest for survival, but acquire an aura: art *makes* human actions special, makes them meaningful and memorable. I would tend to think that, in these perspectives, just as the work of the form would come from a will to resorb the natural asymmetries and roughnesses, the art would be thought as allowing to preserve the social harmony. It would not change anything to the human condition, but would intervene as therapeutic: it would reinforce the beings, would ensure the interior cohesion of the groups, would facilitate our relation to the nature. We find this problematic in Frederick Crews: "Those of us who embrace Darwinian knowledge without cavil are convinced that all existence is unplanned and therefore quite pointless, leaving humanity with the task of *rendering* its life dignified in moral, intellectual, and aesthetic ways scrounged and adjusted from our evolved heritage of repertoires. When the gods have been shipped to fairyland to rejoin the Easter Bunny, we can direct our awe toward beings who actually deserve it — Shakespeare, Rembrandt, Beethoven, Einstein — without cheapening their achievements by ascribing them to mysterious infusions of spirit" (Crews, 2005: xiii). In other words, with the naturalization of the human existence, dies the metaphysical interpretation of the art, but is born, in a suppletive way, a conception of the art as compensation or supplement of sense.

5.2.6. Art as a form of empathy

Another argument could be advanced to explain the advantages of the recourse to the fiction, it is the idea that the art in general, and in particular of the human detour by the representative arts and the fiction favors the empathy, conceived here as mode of social regulation by emotional transfer. In other words, the aptitude for literature would take part of a selection and a valorization of altruism as an indispensable mechanism of self-defense of the species, or simply as fluidification of the social relations. I am thinking here of the often-cited work of psychologists Raymond Mar and Keith Oatley on the way in which exercise through fiction improves our capacity to connect with others (Mar and Oatley, 2005; Mar, Oatley, Hirsh *et al.*, 2006; Mar and Oatley, 2008; Mar, Oatley, and Peterson, 2009), or of Mulligan and Habel's work (2011) on the capacity of fiction to favor the development of intersubjective communication capacities and to fluidify the social behavior of children. It is a question of valuing literature as a moment of constitution of affective communities or as an activation of our capacities for empathy through the power of fictional enunciation to lead us to change our position emotionally and intellectually. An evolutionist like Gottschall combines these two ideas, that of resistance and that of transfer: fiction is an exercise of our prosocial capacities. It "enhances our ability to understand other people; it promotes a deep morality that cuts across religious and political creeds" (Gottschall, 2012b). Gottschall is not far from Butler or the moral theorists of empathetic role change (the theory of role imagination comes from Cavell, in particular), but reinscribes himself in an evolutionary logic where the capacity to adapt to difficulty emphasized by Ellen Dissanayake is essential: literature provides us with both a stimulating and useful access to the interiority of others, but also the illusion of a life that can be lived: "[...] fiction's happy endings seem to warp our sense of reality. They make us believe in a lie: that the world is more just than it actually is. But believing that lie has important effects for society — and it may even help explain why humans tell stories in the first place" (Gottschall, 2012b). Hence, for example, a theorist like William Flesch explains that fictions in the world are dominated by retributive justice or *poetic justice*, by this need to maintain in the optimism of the species (Flesch, 2008).

As some works of anthropology of the literature (for example Thomas Pavel in *The Thought of the novel* that shows that even the individualizing and hyper-speculative reflexivity of the literature possesses a provides "a substantial hypothesis on the nature and the organization of the human world" to value of axiological framework [Pavel, 2003: 46-47]), These works deploy a utilitarian thought of the aesthetic behaviors at the same time stimulating and risky, making of the literature a tool of regulation and social coordination, proposing to make of the literature not only the sharing and the debate of moral values, but the creation of a kind of collective self-regulating instinct.

5.2.7. Pure cognitive gain theories

Another form of functionalist consists not in articulating art to direct stakes of selection or adaptation, but in underlining how, in a more diffuse way, the aptitude of art, and in particular of the narrative forms, to increase our cognitive capacities and thus to support our domination of our environment and the aptitude of the human space to set up and transmit complex social structures. Based on Marie-Laure Ryan's analysis (Ryan, 2010: 482), we can cite various "inflationary" hypotheses that make narrative literature a fundamental mental exercise: all memories take a narrative form (Roger Schank); humanity developed language to respond to a need to tell stories (Mark Turner); narrative constructs reality, identity is a narrative construction and narrative teaches us to read the minds of others (Jerome Bruner); experience is not the subject of narrative, but is made possible by it (David Herman). The general agreement is made around the idea of a mental training: according to Morizot, fiction is a "breaking in of our cognitive capacities" (Morizot, 2012: 185), and to quote Joseph Carroll: "art, music and literature are not merely the products of cognitive fluidity. They are important means by which we cultivate and regulate the complex cognitive machinery on which our more highly developed functions depend" (Carroll, 1998: 481). In other words, because we call in genetics the Baldwin effect, that is to say the integration in the instincts and therefore in the genes of learned cultural behaviors, nature will select the phenotypes of those who do the best learning and will encode in the genotypes aptitudes of cerebral organization optimized by the exercise, passive or active, of the fiction. For this perspective, the human being, in order to assure his domination as an alpha species, is confronted with the necessity of thinking complex and counter-instinctual actions, of managing situations where he overcomes the *stimuli*, of producing long-term contracts, etc. Original behaviors in nature such as technological mastery, post-menopausal survival of females, very long gestation, identification with extended groups could not be thought of without forms of organization of values in time allowed by fiction. As a high-level mental exercise, narrative fiction would favor the mastery of our environment by allowing us to inscribe ourselves in extended logics and imaginative adaptations. These original conducts would impose a kind of "psychological exile" of the man in the nature and would regularly enter in conflict with the determinisms that are the instincts of reproduction and survival. These conflicts between the strategies of the reptilian brain and those, richer, of the neocortex, would be at least as present in the literary representations as the dialectic of the social forces or the relation of the individual to the history.

Many works reference John Tooby and Leda Cosmides' seminal article, "Does Beauty Build Adapted Minds?" (2001). According to Morizot, these authors "developed the idea that in neurocognitive adaptation, the organizational mode plays a role that is at least as fundamental as the functional mode; it is the mode that favors the development of talents, allows their maturation, and tends to signal what would be advantageous to pay attention to outside of instrumental reasons. In short, it is not beauty that is selected, but certain scenarios of response to types of situation acquire an aesthetic significance" (Morizot, 2012: 185). In a rather similar way, for Karl Eibl, the existence of an "induction instinct," an "evolved cognitive tool for information gathering and experiential learning," would explain that rhetoric or poetry, but also the work on formal symmetries (or asymmetries) or the interrogation on the narration and its logics, are indispensable exercises to maintain and select the *fitness* of the individual (Eibl, 2004 and Eibl, 2009, quoted in Mellmann, 2011: 311), and one will find in Changeux (1994; 2008) other analyses going in the same direction and being situated at the crossroads of an archeology of the aesthetic function, reflections stemming from the experimental psychology and the first data stemming from the neurosciences.

H. P. Abbott (2000) or Katja Mellmann (2010) had underlined the intellectual productivity of narrativity and its centrality. The latter places it at the center of “a complex cultural congregation of innate dispositions” (Mellmann, 2011: 313 n. 41), combined with other high-level adaptive capacities, such as those of inducing reasoning and attributing causality to facts. Mark Turner goes further by making narrativity the very origin of language. For him, “parable is the root of the human mind — of thinking, knowing, acting, creating, and plausibly even of speaking” (Turner, 1997: 168). For the American theorist, storytelling is indeed at the center of three fundamental capacities: telling, projecting, and producing a parable. These capacities result from our capacity to put into action what Turner considers as spatial scenarios. Literature enriches this capacity by allowing for shifts between conceptual orders or changes in spatial point of view. It is therefore necessary to analyze the stories as a combination of mental spaces (*blended spaces* [96]), explains Turner (characters, level of reality, etc.). These spatial micro-schemas are at the very origin of our need and our faculties of expression: on the one hand they produce basic grammatical constructions (predication) and on the other hand they provide transpositions (an action can be transposed into another order of spatial schemas). For this micronarrative thought, each sentence is a microhistory. Against Aristotle, but also against Chomsky, Turner refutes the idea that there is a logical structure of language underlying it or a cerebral module built by evolution and dedicated to this capacity[5]: narrativity is not a consequence of the brain’s organization, but its source, a hypothesis that makes the one engaged by literary narrative fiction prevail over any other form of cognitive exercise.

5.3. The heuristics of cognitive psychology in practice

What heuristic model, what type of question is offered to us when evolutionary theory is interested in particular texts, in the works themselves, when we do *close reading*? Evolutionary psychology does not intend to limit itself to the analysis of the “origins” of *Homo narrans* and

---

[5] We are close to the criticism addressed by Quine to the “copy theory of language:” see Quine 1969a: 27.

fossilized behaviors, it is not limited to an archaeology (to what M. S. Sugiyama [2005: 177] calls a "reverse-engineering" functioning of modern fiction), but it wants to show how literature and in particular fiction continues to *respond* to a demand for adaptation by staging concrete adaptive problems, which gives a universal relevance to evolutionary analysis.

It is in thematic terms that such an explanation can be made: for Joseph Carroll, literature produces a "cognitive mapping," an image of human experience colored by subjectivity, a mapping that makes experience intelligible (Carroll, 2011a: 10).[6] This "mapping" analyzes the deep motivations of *Homo sapiens sapiens*: education of the children, efforts of enrichment, competition and cooperation (the fundamental mechanisms of perpetuation and defense of the species) — "the most effective and truest works of literature are those that reference or exemplify these basic facts," comments D. T. Max (2005). In literature, cognitivists argue to produce thematic analyses, character is defined in terms of agency, setting in terms of milieu, and plot in terms of action (Carroll, 2011b: 10). Sugiyama's recent point (2006/2007) gives some idea of the directions taken by literary Darwinists: study of the choice of sexual partners in the Anglo-Saxon novelistic corpus (Sugiyama, 1996; Sugiyama, 1997; Whissell, 1996); highlighting of the reproductive concerns at work in adultery, through the Arthurian narratives (Nesse, 1995)[7]; definition of epic literature as a staging of intrasexual competition, males against males and females against females (Fox, 1995; Fox, 2005; Barash and Barash, 2002; Barash and Barash, 2005; Gottschall, 2001; Gottschall, 2008b); analysis of the behaviors of revolt against the adaptive pressure (Carroll, 2011c); study of jealousy and cuckolding as answers to the competition of the dominants in Pushkin (Cooke, 1999); quantitative approach of the agonistic structure of the Victorian novels (Carroll, Gottschall *et al.*, 2011), etc. It is the whole "life cycle" (Carroll, 2005: 83) and the behavioral activities of man determined by the "hierarchical motivational structure" of its nature (87): an author (i.e. a particular phenotype[8])

---

[6] The notion of cognitive mapping is developed in another text : Carroll, 1995: 3 ff.

[7] The article is devoted to the story of Guinevere and Lancelot in Chrétien de Troyes, Thomas Malory, Tennyson and William Morris.

[8] See Sugiyama 2006/2007: "A given story takes a set of people, each with a different phenotype (i.e., different personality traits, life experiences, fitness attributes and goals), places them in a particular set of historical, cultural,

stages a simulation where fictitious phenotypes are confronted in various environmental contexts with adaptation problems. This system is the same in the character as in the writer and the reader: as Tim Horvath notes, among literary Darwinists, "anything that can be said about authors can by definition also be related to characters in some way, and vice versa" (Horvath, 2005), at the risk of underestimating the distance that can be taken from biological determinisms, as Mellmann suggests, "Literary writers are not *compelled* to fashion verisimilar ('mimetic') representations" (Mellmann, 2011: 311)[9].

Let's take a few specific examples of these analyses, starting with a provocative case: in an essay entitled *Madame Bovary's Ovaries* (2005), David and Nanelle Barash relate the amorous representations of romance literature to a series of adaptive problems: the selection of the best partner, adultery and its stakes in genetic terms for the species, the education of children and the optimal choices of the family. The novelistic logic of Flaubert's biographical novel is, within this explanatory framework, to stage the adaptive optimization strategy of his eponymous character, governed by his ovaries to find a better partner than Charles. Let's give another example: Gottschall on Homer in *The Rape of Troy* defends the idea that the Iliad tells us about the fights of humans returned to their animal nature: "Homer shows that men in combat stoop to a nearly quadrupedal level and concern themselves only with the animal processes of maintaining and destroying life" (Gottschall, 2008b: 162); they are "precocious killer apes who have applied their grand technologies and cooperative instincts to traverse broad seas and rob other men of all they hold precious: their wealth, their women, their lives" (163); they all act unconsciously in accordance with "life's prime directive: be fruitful and multiply" (163). But for Gottschall, Homer also shows how human beings seek to overcome this animality, sometimes succeeding (Achilles returning Hector's body), failing elsewhere (164): in this sense, the epic would be not only a thematization, but a real literary thought experiment about the evolutionary capacities of man. The moral issues often placed at the heart of the novel do not escape such a biologizing deconstruction, as William Flesch's important book

---

and geographical conditions, then plays out one possible version of the interaction of these variables over a certain length of time."

[9] The criticism is brought against Carroll in Richardson, 2000: 561.

Comeuppance (2008), devoted to deserved punishment, costly signals, altruistic punishment and other biological components of fiction, illustrates. Flesch also explains the human attraction to fiction through an evolutionary version of social game theory. The need for fiction comes from the fact that man is a social animal. Therefore, on the one hand, it is a matter of treating narrative as "a verisimilitudinous record of human cooperation" and, on the other hand, of making the work "an object of the kinds of interest that human cooperation requires and rewards" (Flesch, 2008: 182) "Narratives depict signals" (in the sense of evolutionary theory) "and narratives signal" (91). For Flesch, our need for narratives is explained by our need to monitor and appropriate forms of reciprocity: by reading, the reader acquires an aptitude ("fitness") for understanding and exchanging in the social relationship, a capacity and empathy, the latter understood as an indirect reciprocity that makes each individual more interesting and attractive while endowing him or her with a better capacity for social interaction. Being able to know and care about a novel manifests our ability to engage in altruistic activities. This leads to considerations about the value of novels: "We care about the narrative report of what some people do to other people because we care about whether they treat them altruistically or selfishly" (155-156). *Altruistic* here differs from *generous*: it is altruistic to punish the other for their wrongdoing, as it is selfish to neglect the obligation to punish wrongdoers. In sum, we are interested in the novel in what is "prosocial", revenge, for example, as altruistic punishment (Achilles, Hamlet, Batman). Whereas the situation of modern literature, that of an art for art's sake, empowered and breaking with the words and values of the tribe, might seem irreducible to such a functionality, because it places at the center of the literary scene a marginalized individual, a provocative and revolting writer, and the testimony of an irreducible singularity, Flesch very skillfully manages to reintegrate the modern writer, conceived as one who assumes the pain and difficulty of being aware of the world and who shares this ability through an altruistic self-sacrifice (147) — to the point of quoting Leiris and his conception of literature as a bullfight (143). Flesch explains the "aesthetics" of modernity as the complexification and extreme weighting of the signal, and as the place where refined and prosocial forms of cooperation are enacted. In the same spirit and taking up this theory of "deserved punishment," Blakey Vermeule analyzes *Middlemarch*, by George Eliot, showing that the genius of the novelist was to build her novel not on the good feelings and the logics proper to the recognition, but on the moral darkness of the man and his desire to punish the others for their misdeeds (Vermeule, 2011). To generalize the thought of evolutionists, our interest in novels is explained not by a quest for individual morality, but rather by that of collective morality, which is itself overdetermined by the logic of the species, not that of particular individuals. According to a paradox already analyzed by Kant, our anti-social aptitudes are part of our social existence. We see again the similarities of such a theory with the axiological vision of Thomas Pavel in *The Thought of the Novel*, for whom the novel plays the role of a toolbox for solving an axiological problem: the novel "poses above all, and with an unparalleled acuteness, the axiological question of whether the moral ideal is part of the order of the world [...]. In the novel, [...] the axiological question amounts to asking whether, in order to defend the ideal, man must resist the world, immerse himself in it in order to re-establish the moral order, or finally try to remedy his own fragility, if, in other words, the individual can *inhabit* the world in which he is born (Pavel, 2003: 46-47).

Another track of evolutionary analysis of literary works is an analysis of a quantitative nature, illustrating the claimed empiricism of literary Darwinists as well as their verification method. J. Gottschall in *Literature, Science, and a New Humanities* (2008a: 91 ff.)[10] thus starts from an evolutionary hypothesis according to which the female protagonists of the stories should defend their offspring rather than those of others; they should choose men with wealth or high social status and who are physically attractive; they should be less active and less physically heroic than their male counterparts (because, in most sexual species, males have risk-taking and status-seeking behaviors that expose them to the very clear-cut alternative of prolific reproduction and possible death); and finally, readers should prefer active and physically heroic male characters. Gottschall confronts this hypothesis with a corpus of 658 tales from different cultures to draw several conclusions that he believes are convincing: the female heroine role is generally eccentric to the male character's role, they are mostly young and beautiful, they seek to marry and succeed in doing so at the end of the story; they first seek the hero's kindness (prosocial) and they then devote themselves to the good and wealth of their kin, but less through active physical struggle than their male counterparts. He deduces from his statistics empirical universals, which he then uses against feminist analyses that defend on the contrary the cultural and purely Western dimension of our representations of women: the sexual

---

[10] See the review of this book by Easterlin, 2009.

differences examined by crossing their cultural representation are coherent and stable (Gottschall, 2008a: 125) and correspond to what can be expected from the behaviors of sexualized mammals. If the American critic defends himself from being essentialist or anti-feminist, he refutes the culturalist theories of gender as empirically inaccurate, while conceding that the statistically observed differences do not deny the possibility of cultural action and plasticity of the species in the face of biological determinisms.

5.4. Academic and philosophical issues of these emerging disciplines

Such an analysis sufficiently illustrates the stakes of Darwinian interpretations of literature. In the American context of cultural studies, they break with the traditional culturalist explanatory frameworks of criticism and their ideological undertones, since they lead to the reintroduction at the center of reflection of the idea of a human nature, which is certainly evolving and not predetermined. In the French context, the break is made in relation to another dominant critical paradigm, that of formalism, which affirms the autonomy of linguistic structures. The issue at stake is first of all epistemological, for the Darwinians justify their approach by a certain failure of the human sciences in the face of the scientific requirements in which they declare themselves to be inscribed: it would be a question of overcoming the aporias of a literary analysis based on linguistic, cultural or psychoanalytical models and its prohibitive incapacity to become a tool of empirical prediction, despite the structuralist games with white boxes (Gefen, 2006) or Marxist literary teleologies.

The idea of a methodological bankruptcy of the traditional humanities which would refuse experimentation leads Gottschalll to propose to speak about "new humanities" benefiting from the contribution of the life sciences to propose empirical analyses (Gottschall, 2008a), involving in passing virulent reactions like that of the English neuroscientist Raymond Tallis speaking of "*Neurotrash*" about the eliminationist postures of Daniel Dennett and Richard Dawkins (Tallis, 2011b; see also Parry, 2011). In his book *Aping Mankind* (2011a), Tallis designates, not without caricature, two reductionist drifts of the cognitive sciences: on the one hand, *neuromania*, i.e. the idea that brain activity would be the sufficient condition of human consciousness and the only key to our behavior (for example, a famous study of detection of the brain regions involved in love thanks to functional magnetic resonance imaging [Bartels and Zeki, 2000]); on the other hand, *Darwinitis*, i.e. the idea that the theory of evolution would explain not only the origin of the human species (which Tallis admits as a good Darwinian), but also the behavior of man and the nature of social institutions.

Tallis is far from being the only one to brandish the axe of war. To summarize the polemics led notably in *Style* and *Poetic's Today*[11], the criticisms opposed to the evolutionists are made in the name of an autonomist position defending the insularity of the art and the differential character of the artistic practices in the name of the freedom (by making of the art the other of the society and the human animality), of an epistemological skepticism as for the inferences produced with regard to an inaccessible past of the prehistoric man, of reticence in front of the immaturity of the discipline or to its reductionist ambitions. Thus, M.-L. Ryan (2010: 481) insists for example on the speculative character of the theories of the narration as adaptation: "since we do not have at our disposal a specimen of Neanderthal or Cro-Magnon man, we cannot compare the storytelling abilities of mankind during various stages of its cultural or biological evolution, and we can only make educated guesses about the role of storytelling in the development of the social organization typical of humans," explains the Swiss critic. It would be easy, in this respect, to denounce poorly mastered and controversial conceptual transfers, such as, for example, the hypothetical role of mirror neurons in altruistic behavior, extended from macaques and bonobos to evolved human societies (notably by Frans de Waal). This lack of conceptual finesse would render these naturalizing doctrines incapable of expressing particular historical and generic variations[12] without reducing them to invariable universals. Now, for literary Darwinists, for example Mellmann and Carroll, this reduction is inevitable, because it is consubstantial to any real production of knowledge; for them, as for Quine or Dewey, we have nothing to lose by embracing a naturalist and behaviorist conception of signifying facts. The question is not whether Darwinians reduce texts, but how. They do so, Carroll replies, by providing a "*comprehensive* framework" for comparing authors'

---

[11] See especially Eibl and Mellmann, 2008, and *Poetics Today*, 2009. In a more mainstream version, see Gopnik, 2012.

[12] This is, for example, the view of Jackson, 2000: 329, 341; Kelleter, 2007: 164-169. Mellmann, 2011 argues against this assertion. Schaeffer, 2009: 37-43 adopts a measured position.

perspectives, the organization of meaning in texts, and readers' responses (Carroll, 2011b: 29-30; Mellmann, 2011: 310).

The criticisms are not only epistemological: for literary Darwinists, a politically motivated resistance is at work against evolutionary psychology or, if you like, against its scientism and determinism. In the United States, the fight was first waged on the cultural studies front: M. S. Sugiyama (2006/2007) recounts that when she asserted in her dissertation that literary characters were representations of evolved human psychology, and that literary analysis should therefore be founded on an understanding of evolutionary psychology, she was accused of reductionism and took more than a year to form a dissertation jury at the University of California, Santa Barbara. She herself warns against the danger that evolutionary psychology is content to propose a universal interpretative scheme in front of the works, to ritualistically unveil unconscious and deeply buried mechanisms, just like Freudism against which it was built. "Stories do not simply reflect adaptive problems and the cognitive mechanisms that have evolved to solve them. Complex adaptations are facultative: they are sensitive to environmental variation, capable of generating different responses to different environmental inputs. Stories enact the facultative nature of our evolved psychology." (Sugiyama, 2006/2007)[13] It is their utilitarianism, which would be that of Social Darwinism that literary Darwinists have also been reproached with, an accusation of fatalism against which Gottschall defends himself: "evolutionary study of human behavior and psychology is not [...] an ideology of pessimism, defeatism, or conservative defense of the *status quo*" (Gottschall, 2008a: 34); on the contrary, it obeys the Delphic imperative *Know thyself*, the learning of freedom being done in the knowledge of biological determinism: this is the old battle of the culturalists against the biologists, which goes back to the fight of Boas and his students against Darwinism (30 ff.). For the critics of evolutionism, the normative naturalization that this doctrine underlies would hide an inegalitarian right-wing thought. The accusation was made, for example, by Patrick Hogan against Randy Thornhill and Craig Palmer's *A Natural History of Rape* (2000), which presented the propensity of men to commit rape as an evolutionary adaptation (Hogan, 2003b:

---

[13] The author admits to being close to the positions of Easterlin, 2001.

200). Gottschall, taking stock of this debate, pointed out that such hypotheses, which are indeed highly questionable, are the subject of debate within evolutionary psychology *itself*, while Peter Singer defended the existence of a left-wing Darwinism (Gottschall, 2004a: 206; Singer, 2000).

In the face of these criticisms, which explain the still marginal and almost sectarian character of evolutionists in American faculties, intermediate critical positions have been suggested, for example by restricting the ambitions of the method. For D. T. Max, "in the end, literary Darwinism may teach us less about individual books than about the point of literature." (Max, 2005) On the contrary, modern literature would represent problems not reducible to issues of adaptation, and the filter of the evolutionary grid would be too crude to understand literary strategies and denaturalized conduct. For Schaeffer, the animal model is useful, but insufficient to describe human specificities: in the theory of costly signals, the mechanisms of art and courtship are similar, but not the goals and functions. It is necessary according to him to distinguish between structural homology and functional identity, under penalty of engaging in a reductionist approach (Schaeffer, 2009: 37-43). Indeed, "the aesthetic relation is not a function: it is defined as an attentional dynamics regulated by the index of attractiveness of the attentional activity itself" (33).

Another interesting fallback position is represented by the critic Katja Mellmann: for her, evolutionary psychology should not be applied to artifacts such as stories and characters, but to the producer and the reader; it can only concern the living: "anthropological model *reader*" (Mellmann, 2011: 301). Katja Mellmann is in sharp contrast to Carroll and Gottschall's view that problems of adaptation are represented and thematized in literature. For her, it is necessary to work on concrete living subjects and to produce a *stimulus* pattern heuristic specific to reading: "rather than searching for (redundant) analogies between the behavior of fictional persons and actual human beings, we should pay attention to the potential isomorphisms between text structure and innate releasing schemata" (308), because "literary artifacts […] can be said to act as a dummy stimulus on our innate dispositions" (309), that is to say, on our natural receptors of sensation. Such an analysis, rather than leading us to utilitarian speculations, would allow us for example to describe how fiction is "a particular strategy of fictionalisation as based on our adaptations for laughter and humour" (311; refers to Mellmann, 2009: 65-86 and Mellmann, 2007: 264-352, 450-451).

* * *

A few words to conclude and to express both my fascination and my mistrust. As we have seen, this emerging theory, represented by a whole generation of American "literary Darwinists" but also a French school, offers a global explanation of the literary fact that breaks with our whole French critical tradition: for the Darwinians, the literary fact is not marginal in human behaviors, but central to the species; it is not anti-social, but regulating; literature has no ontology of its own or endogenous logic, but is biologically overdetermined; texts or their interpretation are not a space of freedom, but a cognitive device, etc. This vision of literature echoes a new grip of literature on social behaviors, as witnessed by so many essays on personal development and self-management (how X or Y can help you change your life[14]) — which explains the extremely popular character of the works of Alan Palmer, Mark Turner or Jonathan Gottschall, who do not hesitate to explain to us, to quote the title of one of Gottschall's articles, "why fiction is good for us." Certainly, it seems to me that the tools proposed are much more refined than those of neuroscience: While stylistics or cognitive narratology are still in their infancy, and medical investigative work on the aesthetic brain is still in its infancy and unfit for describing the complexity and variety of the effects of literature, the evolutionary paradigm produces powerful and ordinary descriptive categories formulated in a common language, as well as a compelling genealogy of the necessity of that strange and yet so profoundly human-defining (Aristotle's) process that is fiction. The emerging cognitive literary studies

---

[14] See, for example, De Botton, 1997 — a work that admittedly plays the humor card — or Housden, 2001, an anthology whose author has produced other collections with a similar aim: ten poems to say goodbye, to free yourself, to open your heart). Similarly, Méra, 2009 delivers a very serious Balzacian management lesson.

— rhetoric[15], poetics[16], grammar[17], stylistics[18], narratology[19], semiotics[20] — can be built largely from this archaeology of the artistic fact as much as from observational results with obvious benefits for literature: just as neurosciences allow for an objectification of empirical knowledge and functional enlightenment on cognition (neurophysiological bases of reading, cognition, etc.), the evolutionary paradigm can be used as a basis for the development of a new approach, facilitating the reinsertion of literature into common cognitive processes, proposing us in the process to test a new vocabulary and new problems of text analysis. Another interest exemplified by Gottschall's works (in particular 2004a), but also by those, less immediately biologizing, of Patrick Hogan, in particular *The Mind and Its Stories* (Hogan, 2003a; see also Hogan, 2006; Hogan, 2011), it is to relaunch the criticism to the search of human universals by the literature, quest to which we had largely renounced. Thus, in Hogan's affective evolutionism and his theory of the archetypes of the imagination, we could identify "emotional prototypes," based on three different relationships to emotions, universals with a transcultural value. In any case, the confrontation with the Darwinian paradigm engages a truly interdisciplinary debate, whether or not one adheres to Edward O. Wilson's sociobiology affirming the unity of knowledge (Wilson, 1998), one cannot but be fascinated by the hypothesis that "literature and its oral antecedents derive from a uniquely human, species typical disposition for producing and consuming imaginative verbal constructs. Removing the

---

[15] The term "cognitive rhetoric" appeared in the mid-1970s in an article by Dan Sperber (Sperber, 1975). See, among others, Lakoff and Johnson, 1980; Turner and Lakoff, 1989; Turner, 1997.

[16] The term "cognitive poetics" was coined in the early 1990s by Reuven Tsur (Tsur, 1992). See also Stockwell, 2002; Gavins and Steen, 2003; Vandaele and Brône, 2009.

[17] See, for example, Langacker, 1987-1991; Langacker, 1991; Langacker, 2008.

[18] See, for a first look, Semino and Culpeper, 2003; Toolan and Weber, 2005; Burke, 2008.

[19] The movement was launched in the late 1990s by Manfred Jahn in Jahn, 1997. The most recent points are due to David Herman (1999; 2002; 2003; 2007; 2009). Among the founding works, see Cook, 1996; Fludernik, 1996; Emmott, 1997; Abbott, 2001, Richardson and Steen, 2002; Palmer, 2004; Richardson and Spolsky, 2006; Zunshine, 2006.

[20] See in particular Rastier, 1991. Rastier draws on the computational functionalism of Fodor, 1997 and Pylyshyn, 1984; but, for him, this cognitive semantics must be based on what he calls a "material hermeneutics" (see Rastier, 2005).

methodological barrier between humanistic expertise and the expertise of the social sciences can produce results valuable to both fields." (Carroll, Gottschall *et al.*, 2012; cited by Carroll, 2011b: 35) Hence the birth of new disciplines such as the ecocriticism of Harold Fromm or Glen Love (*ecocriticism*) (Love, 2003; Glotfelty and Fromm, 2009; Fromm, 2009), the Darwinian literary history (different from that of the 19th century and Brunetière) of Tom Dolack, based not on the metaphor of the species and the reification of the forms, but on a solid theory of the technological innovation by imitation in the species, etc. (Dolack, 2010). Or Robert Storey's theory of genres and theory of representation in *Mimesis and the Human Animal* (1996; see the review by Carroll, 1996).

With the interdisciplinary postulate of literary evolutionism, it is here the institutional and academic place of the humanities that is questioned. Whether one conceives of literature as an anthropological database, or as a laboratory of experimental psychology or "experimental philosophy," to use another fashionable concept, it is to a relegation of literary criticism and to a reintroduction of the latter into the heart of other knowledge that we could witness. Such are, for example, the virtues of Turner's narrativist theories that I have described and of the cognitivist turn in general: "since the publication in 1980 of George Lakoff and Mark Johnson's influential *Metaphors We Live By*, literary critics have been encouraged by the idea of a cognitive poetics — of, that is, a systematic theory of the mind in which literature is not merely peripheral but central to the understanding of human psychology," writes Van Oort (2003: 238). In other words, literature does not only have to passively import an exogenous conceptual vocabulary, and criticism does not only have to account for its empirical metadiscourses with new scientific tools, but it can serve, differently from any other form of knowledge, to think concrete mental and social processes, by reinscribing itself in the very long duration of human history[21].

References

---

[21] I am in complete agreement with the conclusions of Ryan 2010.

Abbott, H. Porter (2000). “The Evolutionary Origins of the Storied Mind: Modeling the Prehistory of Narrative Consciousness and Its Discontents”. *Narrative*, 8 (3), pp. 247-256.

Abbott, H. Porter (Ed.) (2001). “On the Origins of Fiction: Interdisciplinary Perspectives”, special issue of *SubStance*, 30 (1).

Barash, David P. and Barash, Nanelle R. (2002). “Evolution and Literary Criticism”. *The Chronicle of Higher Education*, October 18, 2002, pp. 7-9.

Barash, David P. and Barash, Nanelle R. (2005). *Madame Bovary’s Ovaries: A Darwinian Look at Literature*. New York, Delacorte Press.

Bartels, Andreas and Zeki, Semir (2000). “The Neural Basis of Romantic Love”. *NeuroReport*, 11 (17), November 27, 2000, pp. 3829-3834; online at: http://www.vislab.ucl.ac.uk/pdf/NeuralBasisOfLove.pdf.

Boyd, Brian (2005). “Evolutionary Theories of Art”. In Gottschall, Jonathan and Wilson, David Sloan (Eds.), *The Literary Animal: Evolution and Nature of the Narrative* (pp. 149-178). Evanston (Ill.), Northwestern University Press.

Brunetière, Ferdinand (1890). *L’Évolution de la critique depuis la Renaissance jusqu’à nos jours*. Paris, Hachette.

Burke, Michael (2008). “How Cognition Can Augment Stylistic Analysis”. *The European Journal of English Studies*, 9 (2), pp. 185-196.

Buss, David (1999). *Evolutionary Psychology: The New Science of the Mind*. Boston, Allyn & Bacon.

Carroll, Joseph (1995). *Evolution and Literary Theory*. Columbia, University of Missouri Press.

Carroll, Joseph (1996). Review of Robert F. Storey, *Mimesis and the Human Animal*, 1996; online at http://cogweb.ucla.edu/Abstracts/Carroll_S98.html.

Carroll, Joseph (1998). “Steven Pinker’s Cheesecake for the Mind” (Review of Steven Pinker, *How the Mind Works*, 1997). *Philosophy and Literature,* 22 (2), p. 478-485; online at: http://cogweb.ucla.edu/Abstracts/Carroll_C98.html.

Carroll, Joseph (2005). “Human Nature and Literary Meaning: A Theoretical Model Illustrated with a Critique of *Pride and Prejudice*”. In Gottschall, Jonathan and Wilson, David

Sloan (Eds.), *The Literary Animal: Evolution and Nature of the Narrative* (pp. 76-116). Evanston (Ill.), Northwestern University Press.

Carroll, Joseph (2011a). *Reading Human Nature: Literary Darwinism in Theory and Practice*. Albany (N.Y.), SUNY Press.

Carroll, Joseph (2011b). "An Evolutionary Paradigm for Literary Study, with Two Sequels". In Carroll, Joseph, *Reading Human Nature: Literary Darwinism in Theory and Practice* (pp. 3-54). Albany (N.Y.), SUNY Press.

Carroll, Joseph (2011c). "The Cuckoo's History: Human Nature in *Wuthering Heights*" In Carroll, Joseph, *Reading Human Nature: Literary Darwinism in Theory and Practice* (pp. 109-122). Albany (N.Y.), SUNY Press.

Carroll, Joseph, Gottschall, Jonathan, Johnson, John A., and Kruger, Daniel (2011). "Agonistic Structure in Victorian Novels: Doing the Math". In Carroll, Joseph, *Reading Human Nature: Literary Darwinism in Theory and Practice* (pp. 151-175). Albany (N.Y.), SUNY Press.

Carroll, Joseph, Gottschall, Jonathan, Johnson, John A., and Kruger, Daniel (2012). *Graphing Jane Austen: The Evolutionary Basis of Literary Meaning*. New York, Palgrave Macmillan.

Changeux, Jean-Pierre (1994). *Raison et plaisir*. Paris, Odile Jacob.

Changeux, Jean-Pierre (2008). *Du vrai, du beau, du bien*. Paris, Odile Jacob.

Cohn, Dorrit (1978). *Transparent Minds: Narrative Modes for Presenting Consciousness in Fiction*. Princeton, Princeton University Press.

Compagnon, Antoine. "Genre, création, évolution" (Course); online at: http://www.fabula.org/compagnon/genre12.php.

Cook, Guy (1994). *Discourse and Literature*. Oxford, Oxford University Press.

Cooke, Brett (1999). "Sexual Property in Pushkin's 'The Snowstorm': A Darwinist Perspective". In Cooke, Brett and Turner, Frederick (Eds.), *Biopoetics: Evolutionary Explorations in the Arts* (pp. 175-204). Lexington, ICUS.

Crews, Frederick (2005). “Foreword from the Literary Side”. In Gottschall, Jonathan and Wilson, David Sloan (Eds.), *The Literary Animal: Evolution and Nature of the Narrative* (pp. xiii-xv). Evanston (Ill.), Northwestern University Press.

Dawkins, Richard (1976). *The Selfish Gene*. Oxford, Oxford University Press.

Dawkins, Richard (1982). *The Extended Phenotype*. Oxford, Oxford University Press.

Dawkins, Richard (2006). *The God Delusion*. London, Bantam Press.

De Botton, Alain (1997). *How Proust Can Change Your Life*. New York, Pantheon.

Deacon, Terrence W. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. New York, W. W. Norton & Company.

Dennett, Daniel C. (1991). *Consciousness Explained*. Boston, Little, Brown.

Dennett, Daniel C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York, Simon & Schuster.

Dennett, Daniel C. (2006). *Breaking the Spell: Religion as a Natural Phenomenon*. London, Allen Lane.

Dissanayake, Ellen (1992). *Homo Aestheticus: Where Art Comes from and Why*. New York, Free Press.

Dolack, Tom (2010). “Imitation and Literary Evolution”. *Politics and Culture*, (1), April 29, 2010, online at: http://www.politicsandculture.org/2010/04/29/imitation-and-literary-evolution/.

Easterlin, Nancy (2001). “Hans Christian Andersen Fish Out of Water”. *Philosophy and Literature,* 25 (2), pp. 251-277.

Easterlin, Nancy (2009). “Literature, Science, and the New Humanities” (Review). *Philosophy and Literature*, 33 (1), pp. 230-233.

Eibl, Karl (2004). *Animal Poeta. Bausteine der biologischen Kultur- und Literaturtheorie*. Paderborn, Mentis.

Eibl, Karl (2009). “The Induction Instinct: Evolution and Poetic Application of a Cognitive Tool”. *Studies in the Literary Imagination*, 42 (2), pp. 43-60.

Eibl, Karl and Mellmann, Katja (2008). “Misleading Alternatives”. *Style*, 42 (2-3), pp. 166-171.

Emmott, Catherine (1997). *Narrative Comprehension: A Discourse Perspective*. Oxford, Oxford University Press.

Flesch, William (2008). *Comeuppance: Costly Signaling, Altruistic Punishment, and Other Biological Components of Fiction*. Cambridge (Mass.), Harvard University Press.

Fludernik, Monika (1996). *Towards a “Natural” Narratology*. London, Routledge.

Fodor, J. A. (1987). *Psychosemantics*. Cambridge (Mass.), M.I.T. Press.

Fox, Robin (1995). “Sexual Conflict in the Epics”. *Human Nature*, 6 (2), pp. 135-144.

Fox, Robin (2005). “Male Bonding in the Epics and Romances”. In Gottschall, Jonathan and Wilson, David Sloan (Eds.), *The Literary Animal: Evolution and Nature of the Narrative* (pp. 126-144). Evanston (Ill.), Northwestern University Press.

Fromm, Harold (2010). *The Nature of Being Human: From Environmentalism to Consciousness*. Baltimore, Johns Hopkins University Press.

Gavins, Joanna and Steen, Gerard (Eds.) (2003). *Cognitive Poetics in Practice*. London, Routledge.

Gefen, Alexandre (2006). “Du bon usage des tableaux à double entrée: stratégies taxinomiques et ambitions épistémologiques de la théorie littéraire”. In Escola, Marc and Rabau, Sophie (Eds.), *La Case blanche: théorie littéraire et textes possibles*, actes du colloque d’Oléron (14-18 avril 2003). Reims, *La Lecture littéraire*, n. 8.

Gefen, Alexandre (2007). “L’adieu aux exemples. Sérendipité et inexemplarité dans la littérature modern”. In Bouju, Emmanuel, Gefen, Alexandre, Hautcœur, Guiomar, and Macé, Marielle (Eds.), *Littérature et exemplarité* (pp. 249-261). Rennes, PUR.

Gefen, Alexandre (2010). “L’éthique est-elle un récit? Le récit est-il une éthique? Retour sur la querelle du ‘narrativisme’”. Proceedings of the international colloquium “Les moralistes modernes”, University of Belgrade (22-23 October 2009). Published online at: http://www.fabula.org/colloques/document1352.php.

Glotfelty, Cheryll and Fromm, Harold (Eds.) (2009). *The Ecocriticism Reader: Landmarks in Literary Ecology*. 2nd ed.: Athens, The University of Georgia Press.

Gopnik, Adam (2012). “Can science explain why we like stories?” *The New Yorker*; online at: http://www.newyorker.com/online/blogs/books/2012/05/can-science-explain-why-we-tell-stories.html.

Gottschall, Jonathan (2001). “Homer’s Human Animal: Ritual Combat in the *Iliad*”. *Philosophy and Literature*, 25 (2), pp. 278-294.

Gottschall, Jonathan (2004b). Review of Joseph Carroll, *Literary Darwinism*, 2004. *Human Ethology Bulletin*, 19 (4), pp. 11-13.

Gottschall, Jonathan (2008a). *Literature, Science, and a New Humanities*. New York, Palgrave Macmillan.

Gottschall, Jonathan (2008b). *The Rape of Troy: Evolution, Violence, and the World of Homer*. Cambridge, New York, Cambridge University Press.

Gottschall, Jonathan (2012a). *The Storytelling Animal: How Stories Makes Us Human*. New York, Houghton Mifflin Harcourt.

Gottschall, Jonathan (2012b). “Why Fiction Is Good for You”. *The Boston Globe*, April 29, 2012; online at: http://articles.boston.com/2012-04-29/ideas/31417849_1_fiction-morality-happy-endings.

Guerin, Wilfred L., Labor, Earle, Morgan, Lee, Reesman, Jeanne C., and Willingham, John R. (2010). *A Handbook of Critical Approaches to Literature*. 6th ed.: Oxford and New York, Oxford University Press.

Herman, David (Ed.) (1999). *Narratologies: New Perspectives on Narrative Analysis*. Columbus, Ohio State University Press.

Herman, David (2002). *Story Logic: Problems and Possibilities of Narrative*. Lincoln (Neb.), University of Nebraska Press.

Herman, David (2003). “Stories as a Tool for Thinking”. In Herman, David (Ed.), *Narrative Theory and the Cognitive Sciences* (pp. 163-192). Stanford, CSLI.

Herman, David (2007). “Storytelling and the Sciences of the Mind: Cognitive Narratology, Discursive Psychology, and Narratives in Face-to-Face Interaction”. *Narrative*, 15, pp. 306-334.

Herman, David (2009). *Basic Elements of Narrative*. Oxford, Wiley-Blackwell.

Hogan, Patrick Colm (2003a). *The Mind and Its Stories: Narrative Universals and Human Emotion*. New York, Cambridge University Press.

Hogan, Patrick Colm (2003b). *Cognitive Science, Literature, and the Arts: A Guide for Humanists*. London, Routledge.

Hogan, Patrick Colm (2006). "A Debate for Making Clear the Worth of Universals: Ninety-Five Theses". Position paper serving as the text for debate at the plenary session on "Aesthetic Universals". Conference of the International Society for the Empirical Study of Literature and Media (Internationale Gesellschaft für Empirische Literaturwissenschaft). Fraueninsel, Germany, August 2006.

Hogan, Patrick Colm (2011). *Affective Narratology: The Emotional Structure of Stories*. Lincoln (Neb.), University of Nebraska Press.

Horvath, Tim (2005). "Literary Darwinism and Literary Darwinisms: Let the Games Begin…". *Cogweb*, 1 January 2005; online at: http://cogweb.ucla.edu/Culture/Horvath-on-Carroll-04.html.

Housden, Roger (2001). *Ten Poems to Change Your Life*. New York, Harmony.

Jackson, Tony (2000). "Questioning Interdisciplinarity: Cognitive Science, Evolutionary Psychology, and Literary Criticism". *Poetics Today*, 21 (2), pp. 319-347.

Jahn, Manfred (1997). "Frames, Preferences, and the Reading of Third-Person Narratives: Toward a Cognitive Narratology". *Poetics Today*, 18 (4), pp. 441-468.

Kelleter, Frank (2007). "A Tale of Two Natures: Worried Reflections on the Study of Literature and Culture in an Age of Neuroscience and Neo-Darwinism". *Journal of Literary Theory*, 1 (1), pp. 153-189, online at: http://www.jltonline.de/index.php/articles/article/view/65/258.

Lageira, Jacinto (2012). "Artialisation". In Pouivet, Roger et Morizot, Jacques (Eds.), *Dictionnaire de philosophie esthétique* (pp. 49-50). 2nd ed.: Paris, Armand Colin.

Lakoff, George and Johnson, Mark (1980). *Metaphors We Live By*. Chicago, University of Chicago Press, 1980.

Langacker, Ronald W. (1987-1991). *Foundations of Cognitive Grammar*. Stanford, Stanford University Press, 2 vols.

Langacker, Ronald W. (1991). *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin and New York, Mouton de Gruyter.

Langacker, Ronald W. (2008). *Cognitive Grammar: A Basic Introduction*. New York, Oxford University Press.

Laugier, Sandra (1992). *L'Apprentissage de l'obvie: l'anthropologie logique de Quine*. Paris, J. Vrin.

Laugier, Sandra (2010).*Wittgenstein, le mythe de l'inexpressivité*. Paris, J. Vrin.

Love, Glen (2003). *Practical Ecocriticism: Literature, Biology, and the Environment*. Charlottesville: University of Virginia Press.

Mar, Raymond and Oatley, Keith (2005). "Evolutionary Pre-adaptation and the Idea of Character in Fiction". *Journal of Cultural and Evolutionary Psychology*, 3 (2), pp. 179-194.

Mar, Raymond and Oatley, Keith (2008). "The Function of Fiction Is the Abstraction and Simulation of Social Experience". *Perspectives on Psychological Science*, 3 (3), pp. 173-192.

Mar, Raymond, Oatley, Keith, Hirsh, Jonathan, dela Paz, Jennifer, and Peterson, Jordan B. (2006). "Bookworms versus Nerds: Exposure to Fiction versus Non-fiction, Divergent Associations with Social Ability, and the Simulation of Fictional Social Worlds". *Journal of Research in Personality*, 40, pp. 694-712.

Mar, Raymond, Oatley, Keith, and Peterson, Jordan B. (2009). "Exploring the Link between Reading Fiction and Empathy: Ruling out Individual Differences and Examining Outcomes". *Communications, 34,* pp. 407-428.

Max, D. T. (2005). "The Literary Darwinists". *The New York Times Magazine*, November 6, 2005; online at: http://www.nytimes.com/2005/11/06/magazine/06darwin.html?pagewanted=all.

Mellmann, Katja (2007). *Emotionalisierung. Von der Nebenstundenpoesie zum Buch als Freund*. Paderborn, Mentis.

Mellmann, Katja (2009). "Das 'Spielgesicht' als poetisches Verfahren. Elemente einer verhaltensbasierten Fiktionalitätstheorie". In Anz, Thomas and Kaulen, Heinrich (Eds.), *Literatur als Spiel. Evolutionsbiologische, ästhetische und pädagogische Aspekte* (pp. 65-86). Berlin, de Gruyter.

Mellmann, Katja (2010). “Voice and Perception. An Evolutionary Approach to the Basic Functions of Narrative”. In Aldama, Frederick Luis (Ed.), *Toward a Cognitive Theory of Narrative Acts* (pp. 119-140). Austin, University of Texas Press.

Mellmann, Katja (2011). “Evolutionary Psychology as a Heuristic in Literary Studies”. In Saul, Nicholas and James, Simon J. (Eds.), *The Evolution of Literature: Legacies of Darwin in European Cultures* (pp. 299-317). Amsterdam, Rodopi.

Méra, Brigitte (2009). *La Méthode Rastignac: “La Comédie humaine”, une culture d’entreprise*. Paris, Tallandier.

Miller, Geoffrey (2000). *The Eating Mind: How Sexual Choice Shaped the Evolution of Human Nature*. New York, Doubleday Books.

Mithen, Steve J. (1996). *The Prehistory of the Mind: A Search for the Origins of Art, Religion, and Science*. London, Thames and Hudson.

Morizot, Jacques (2012). “Évolution et esthétique”. In Pouivet, Roger et Morizot, Jacques (Eds.), *Dictionnaire de philosophie esthétique* (pp. 184-186). 2nd ed.: Paris, Armand Colin.

Mulligan, Kenneth and Habel, Philip (2011). “An Experimental Test of the Effects of Fictional Framing on Attitudes”. *Social Science Quarterly*, 92 (1), pp. 79-99.

Nesse, Margaret (1995). “Guivenere’s Choice”. *Human Nature*, 6 (2), pp. 145-163.

Nettle, David (2005). “What Happens in *Hamlet*? Exploring the Psychological Foundations of Drama”. In Gottschall, Jonathan and Wilson, David Sloan (Eds.), *The Literary Animal: Evolution and Nature of the Narrative* (pp. 56-75). Evanston (Ill.), Northwestern University Press.

Palmer, Alan (2004). *Fictional Minds*. Lincoln, University of Nebraska Press.

Parry, Mark (2011). “Raymond Tallis Takes Out ‘Neurotrash’”. *The Chronicle Review*, October 9, 2011; published online: http://chronicle.com/article/Raymond-Tallis-Takes-Out-the/129279/.

Pavel, Thomas (2003). *La Pensée du roman*. Paris, Gallimard.

Pinker, Steven (1997). *How the Mind Works*. New York, W. W. Norton.

Pinker, Steven (1997). “Toward a Consilient Study of Literature”. *Philosophy and Literature*, 31 (1), pp. 161-177.

Poetics Today (2009). “Cognitive Themes”. *Poetics Today*, 30 (3).

Prince, Gerald (2008). “Narratologie Classique et narratologie post-classique;” published online: http://www.vox-poetica.org/t/articles/prince.html.

Pylyshyn, Z. (1984). *Computation and Cognition*. Cambridge (Mass.), MIT Press.

Quine, W. V. (1969a). “Ontological Relativity”. In Quine, W. V., *Ontological Relativity and Other Essays* (pp. 26-68). New York, Columbia University Press.

Quine, W. V. (1969b). “Epistelomogy Naturalized”. In Quine, W. V., *Ontological Relativity and Other Essays* (pp. 69-90). New York, Columbia University Press.

Rastier, François (1991). *Sémantique et recherches cognitives*. Paris, PUF. 2nd ed.: 2001.

Rastier, François (2005). “Sémiotique du cognitivisme et sémantique cognitive: questions d’histoire et d’epistémologie”. *Texto*; online at: http://www.revue-texto.net/inedits/rastier/rastier_semantique-cognitive.html.

Richardson, Alan (2000). “Rethinking Romantic Incest: Human Universals, Literary Representation, and the Biology of Mind”. *New Literary History*, 31 (3), pp. 553-572.

Richardson, Alan and Spolsky, Ellen (Eds.) (2006). *The Work of Fiction: Cognition, Culture, and Complexity*. Aldershot, Ashgate.

Richardson, Alan and Steen, Francis F. (Eds.) (2002). “Literature and the Cognitive Revolution”. *Poetics Today*, special issue, 23 (1).

Ryan, Marie-Laure (2010). “Narratology and Cognitive Science: A Problematic Relation”. *Style*, 44 (4), pp. 469-495.

Schaeffer, Jean-Marie (2007). *La Fin de l’exception humaine*. Paris, Gallimard.

Schaeffer, Jean-Marie (2009). *Théorie des signaux coûteux: esthétique et art*. Rimouski (Québec), Tangence éditeur.

Schaeffer, Jean-Marie (2012). “Pour une esthétique naturaliste: l’animal artiste et esthète? “. In Pouivet, Roger et Morizot, Jacques (Eds.), *Dictionnaire de philosophie esthétique* (pp. 28-31). 2nd ed.: Paris, Armand Colin.

Semino, Elena and Culpeper, Jonathan (Eds.) (2003). *Cognitive Stylistics: Language and Cognition in Text Analysis*. Amsterdam and Philadelphia: John Benjamins.

Singer, Peter (2000). *A Darwinian Left: Politics, Evolution, and Cooperation*. New Haven, Yale University Press.

Sperber, Dan (1975). “Rudiments of Cognitive Rhetoric”. *Poetics*, (23), pp. 389-415.

Stockwell, Peter (2002). *Cognitive Poetics: An Introduction*. London, Routledge.

Storey, Robert F. (1996). *Mimesis and the Human Animal: On the Biogenetic Foundations of Literary Representation*. Evanston (Ill.), Northwestern University Press.

Sugiyama, Michelle Scalise (1996). “On the Origins of Narrative: Storyteller Bias as a Fitness-Enhancing Strategy”. *Human Nature*, 7 (7), pp. 403-425.

Sugiyama, Michelle Scalise (1997). *Feminine Nature: An Evolutionary Analysis of Hemingway’s Women Characters* (PhD). University of California, Santa Barbara.

Sugiyama, Michelle Scalise (2005). “Reverse Engineering Narrative”. In Gottschall, Jonathan and Wilson, David Sloan (Eds.), *The Literary Animal: Evolution and Nature of the Narrative* (pp. 177-195). Evanston (Ill.), Northwestern University Press.

Sugiyama, Michelle Scalise (2006/2007). “The Nature of Literature”, review of David and Nanelle Barash, *Madame Bovary’s Ovaries*. *Entelechy*, (8); online at: http://www.entelechyjournal.com/michellescalisesugiyama.html.

Tallis, Raymond (2011a). *Aping Mankind: Neuromania, Darwinitis, and the Misrepresentation of Humanity*. Durham, Acumen.

Tallis, Raymond (2011b). *‘*Neurotrash. Humans Are Special”. *Prospect*, published online: http://www.prospectmagazine.co.uk/magazine/neurotrash-brain-chemistry-biologism-neurones-darwin/.

Thornhill, Randy and Palmer, Craig T. (2000). *A Natural History of Rape: Biological Bases of Sexual Coercion*. Cambridge (Mass.), MIT Press.

Tooby, John and Cosmides, Leda (2001). “Does Beauty Build Adapted Minds? Toward an Evolutionary Theory of Aesthetics, Fiction and the Arts”. *SubStance*, (94/95), pp. 6-27; online at: http://www.psych.ucsb.edu/research/cep/papers/beauty01.pdf.

Toolan, Michael and Weber, Jean-Jacques (Eds.) (2005). *The Cognitive Turn: Papers in Cognitive Literary Studies*, *European Journal of English Studies*, 9 (2).

Tsur, Reuven (1992). *Toward a Theory of Cognitive Poetics*. Amsterdam, North-Holland, 1992. 2nd ed.: Brighton and Portland, Sussex Academic Press.

Turner, Mark (1997). *The Literary Mind: The Origins of Thought and Language*. New York, Oxford University Press.

Turner, Mark and Lakoff, George (1989). *More Than Cool Reason: A Field Guide to Poetic Metaphor*. Chicago, University of Chicago Press.

Van Oort, Richard (2003). “Cognitive Science and the Problem of Representation”. *Poetics Today*, 24 (2), pp. 237-295.

Vandaele, Jeroen and Brône, Geert (Eds.) (2009). *Cognitive Poetics: Goals, Gains and Gaps*. Berlin, Mouton de Gruyter.

Vermeule, Blakey (2011). “A Comeuppance Theory of Narrative and Emotions”. *Poetics Today*, 32 (2), pp. 235-253.

Whissell, Cynthia M. (1996). “Mate Selection in Popular Women’s Fiction”. *Human Nature*, 7 (4), pp. 427-447.

Wilson, Edward O. (1998). *Consilience: The Unity of Knowledge*. New York: Alfred A. Knopf.

Zunshine, Lisa (2006). *Why We Read Fiction: Theory of Mind and the Novel*. Columbus, Ohio State University Press; online at: https://uky.academia.edu/LisaZunshine.

Zunshine, Lisa (2008). “Theory of Mind and Fictions of Embodied Transparency”. *Narrative*, 16 (1), pp. 65-92; online at: https://uky.academia.edu/LisaZunshine.

Zunshine, Lisa (2011). “Style Brings in Mental States”. *Style*, 45 (2), pp. 349-356; online at: https://uky.academia.edu/LisaZunshine.

Chapter 6
Evolutionary Aspects of Language Change

Johann-Mattis List[i]

Abstract

While it has been known for a long time that human languages can change in various ways, it was only in the early 19th century that scholars realized that certain aspects of language change proceed in a surprisingly regular manner, allowing us to reconstruct historical stages of languages which have never been documented in written sources. The findings led to the establishment of historical linguistics as a scientific discipline, devoted to the investigation of how languages change and why. Although evolutionary thinking plays a major role in historical linguistics, practitioners often have the tendency to emphasize the peculiarities of language evolution rather than the commonalities with other kinds of evolution. In part, this seems to be justified by some phenomena for which it is difficult to find counterparts in different disciplines. In part, however, this may also due to a communication problem that is characteristic for interdisciplinary research, since scholars lack a common terminology. As a result, it is difficult for linguists to explain their particular evolutionary views on language change to practitioners from other disciplines, while evolutionary terminology from disciplines such as biology is difficult to grasp for linguists. In the study, I will try to present some important evolutionary aspects of language change for which it is hard to find counterparts in other disciplines and then point to current challenges of evolutionary studies in historical linguistics which have to deal with these aspects.

Keywords: language change, historical linguistics, language evolution, history of linguistics

---

[i] Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology; mattis_list@eva.mpg.de

6.1. Introduction

Language change and language evolution have received much attention of late. Inspired by quantitative approaches from evolutionary biology, scholars have started to pay increased attention to the phylogenetic development of the world's larger language families (Gray et al. 2009; Bouckaert et al. 2012; Bouckaert et al. 2018; Kolipakam et al. 2018; Sagart et al. 2019; Gerardi et al. 2021). Had phylogenetic reconstruction been largely ignored in qualitative research, or reduced to the subgrouping of the major branches of a language family by means of qualitative cladistic analysis, it has now become a popular research topic in historical linguistics and – despite initial skepticism – the majority of scholars now seems to accept phylogenetic studies based on Bayesian inference or maximum likelihood as a powerful set of tools that are valid for the exploration of a language family's past.

The introduction of phylogenetic methods in historical linguistics is accompanied by an increasing amount of discussions devoted to the nature of language evolution in comparison with other kinds of evolution. Following a long tradition of skepticism towards evolutionary explanations of language change, many scholars still emphasize the peculiarity of language evolution in comparison with biological evolution, and some scholars even find it misleading to discuss language change as an evolutionary phenomenon at all. On the other hand, there is a growing number of attempts to model the dynamics of language evolution formally and computationally. In the following, I will try to give a short background on the history of evolutionary thinking in the field of historical linguistics from the beginning of the 19th century until today (§ 2) and then point to four aspects of language change which I consider crucial, in so far as they reflect differences in the evolutionary processes which often do not have a direct counterpart in evolutionary biology. I will then discuss three unsolved problems in historical linguistics for which no solutions have been proposed so far, also due to the fact that they are peculiar for language change. While I am not able to propose a proper solution for these problems, I conclude that future efforts in linguistics should try to concentrate on the adaptation of methods from other disciplines to linguistic needs rather than to the direct transfer.

6.2. Background

The discipline of comparative linguistics has a long tradition of evolutionary thinking, reaching much deeper back in time than that of biology (List et al. 2016). In contrast to biologists, who had to infer that observed biological diversity was the result of a long succession of changes, linguists could observe these changes almost directly through the comparison of documents written in the same language at different times. Having almost direct access to ancestral stages of contemporary languages was a striking advantage of linguistics over biology which was already observed by August Schleicher (1821-1868) in 1863 (Schleicher 1863). It helped linguists not only in the development of methods for the inference of phylogenetic relationships but also allowed them to propose techniques by which the supposed pronunciation of individual words in ancestral languages could be estimated. This technique, known as "linguistic reconstruction" (Fox 1995) is still the key objective of historical language comparison and linguists consider it as as much more important than the reconstruction of phylogenies.

6.2.1. From Words to Trees

Linguists have known for a long time that languages evolve and that the languages we observe today may stem from common sources which themselves no longer exists. First speculations on the common descent and the tree- or network-like separation of languages can already be found in early studies of the 17th century and thereafter (List et al. 2016). Until the late 18th century, however, the dominant view among scholars in Europe was that all human languages were products of the mythical Confusion of Tongues which prevented the construction of the Tower of Babel (Klein 2004).

First systematic investigations in which languages were compared for their genetic relatedness were carried out towards the end of the 18th century and reached first popularity with the detection of the Indo-European language family, represented specifically by Sanskrit, Ancient Greek, Latin, and Gothic (Rask 1818; Grimm 1822), which was later expanded by more languages from other branches (Slavic, Albanian, Armenian, see e.g. Hübschmann 1877), and reached its current state with the detection of Hittite (Hrozný 1915).

Unlike modern phylogenetic trees, early linguistic trees were much less formal and systematic, but had the tendency to resemble true trees much more closely. As an example, consider Schleicher's tree from 1853 (Schleicher 1853), which has the appearance of a massive oak with a big trunk. Only later, the family tree visualizations became more schematized, but the interpretation was still far away from being formalized.

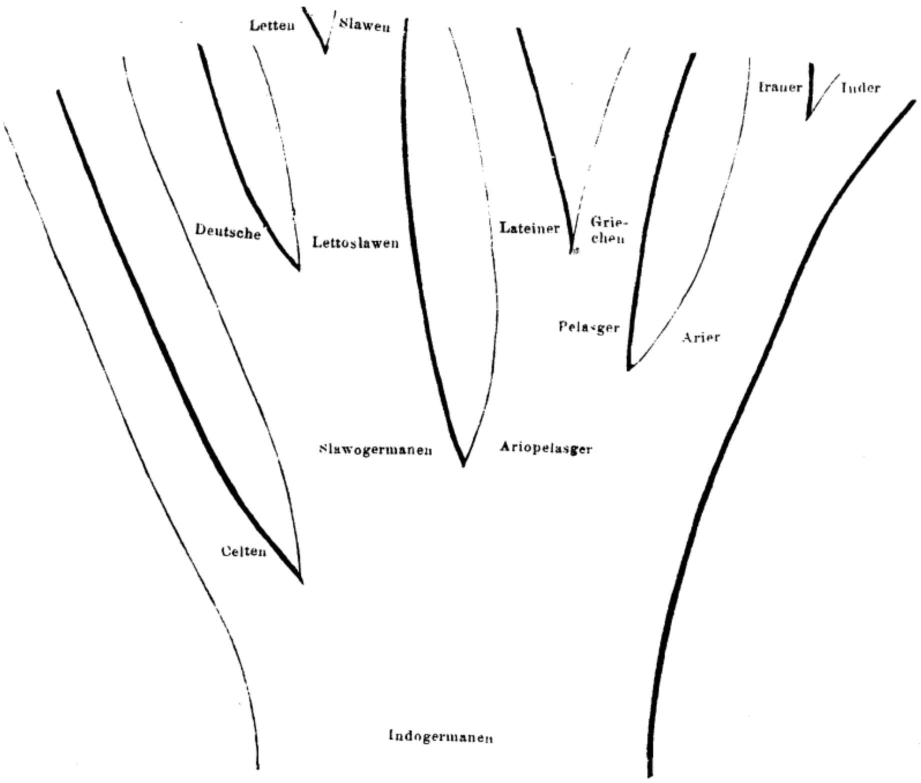


Figure 1*: August Schleicher's Oak Family Tree from 1853.*

As an example for the lack of formalization, consider again a tree by Schleicher, this time from 1861 (Schleicher 1861). While this tree looks much more formalized than the earlier tree from 1853, the description of this tree in the text is interesting, since Schleicher points to branch lengths as representing the supposed time which had elapsed since separation while at the same time emphasizing that the distance between extant languages reflected their synchronic closeness. While the German passage remains unclear in the wording, one way to read it is to assume that Schleicher made a direct distinction between the "closeness" of languages as shown by their horizontal arrangement on a tree and the closeness as derived from the history reflected in divergence times. While it is impossible to depict the former systematically in a two-dimensional drawing, it is possible that Schleicher thought of some additional closeness between languages independent of their evolutionary history and tried to mark this in his tree drawing by separating the major subgroups visually from each other in the tree and by placing languages like Albanian and Greek horizontally close to each other while at the same time assigning them a larger divergence time than given for Celtic and Italian.

> Die ältesten teilungen des indogermanischen bis zum entstehen der grundsprachen der den sprachstamm bildenden sprachfamilien laßen sich durch folgendes schema anschaulich machen. Die länge der linien deutet die zeitdauer an, die *entfernung derselben von einander den verwantschaftsgrad.*
>
> (Translation): The oldest splits of Indo-European until the development of the fundamental languages of the language families which constitute the stem of the language [sprachstamm] can be visualized by the following schema. The length of the lines indicates the elapsed time, the *distance of the lines from each other indicates the degree of relationship.* (Schleicher 1861: 6f, my translation and emphasis)

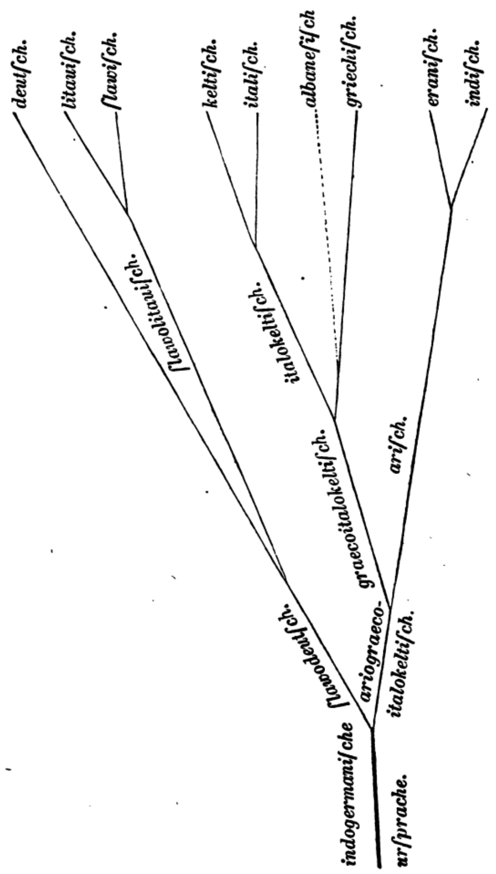


***Figure 2**: Schleicher's 1861 tree, which groups the languge family into tree major groups.*

Even if my attempt to interpret the peculiarities of Schleicher's family tree from 1861 along with his explanations turns out to be wrong: what seems important about this early phase of tree thinking in historical linguistics is that scholars did not use the tree model as a clear-cut tool for mathematical modeling. Instead they used the idea of a branching tree as a source of inspiration for the modeling of phenomena which they could observe but not yet fully understand.

6.2.2. From Trees to Waves

Not long after Schleicher and some colleagues had propagated their family tree models for the first time, scholars began to contest them. One of the most prominently cited opponents of Schleicher's family trees was Johannes Schmidt (1843-1901), who devoted a complete booklet to contradict Schleicher (Schmidt 1872). In this study, Schmidt presented concrete data in the form of sets of homologous words ("cognate sets" in linguistic terminology) for the major Indo-European branches known by then and classified them according to their distribution across the branches. He noted that one could easily find examples for homologs shared exclusively among different possible pairings (Greek vs. Old Indian, Greek vs. Slavic, Slavic vs. Old Indian) with no residues ("reflexes" in linguistic terminology) in any of the other branches. Based on this finding, Schmidt refuted the family tree hypothesis, arguing that a tree could not explain the observed data. What Schmidt proposed instead was the rather fuzzy idea of a wave-like expansion of the major branches of the Indo-European languages which contributed to their gradual separation and would explain the specific commonalities between individual pairs which seemed to contradict each other.

Unfortunately, Schmidt did not see that the cases he listed could be perfectly explained by the traditional tree model assuming well-known phenomena like differential loss (Geisler and List 2013) or incomplete lineage sorting (List et al. 2016; Jacques and List 2019). But although his critic was not valid and his alternative model, the "wave theory" (*Wellentheorie*), as it was called thereafter, did not offer any concrete instructions with respect to the formal modeling of language divergence and spread, many linguists started to present it as a valid alternative to the family tree model and even today, no textbook on historical linguistics can get away to present the family tree model without mentioning the wave theory as an alternative to the tree.

Nowadays, the wave theory is often presented as some kind of diffusion model in which languages gradually diverge without splitting abruptly. This model of language evolution is a valid way to describe language divergence which was already mentioned by Hugo Schuchardt

[Schuchardt (1900), 1842-1927), and it is clear that family tree models are not capable of modeling the split process in detail. Attributing diffusion models to Schmidt's wave theory, however, does not seem to be justified and has led to a lot of confusion among students of historical linguistics (see Jacques and List 2019 for a detailed discussion on this topic).

6.2.3. From Waves to Forests

After the heated debates about trees and waves during the end and the beginning of the 19th century, neither the tree model nor the wave theory were essentially advanced any further. Both models were only described and discussed in prose with no attempts of providing a formal modeling and trees in particular were viewed with a large amount of skepticism. As a result, the tree model needed a very long time to recover: only a few studies in the 20th century explicitly used family trees, and if they did so, they typically contrasted them with the wave model – or what scholars thought the wave model represented. Scholars typically labeled family trees as unrealistic and not capable of depicting language history in all its complexity.

The bad reputation of the tree model is nicely reflected in the work of Morris Swadesh (1909-1967), one of the founding fathers of the modern lexicostatistical techniques of data annotation and data sampling which are still used in modern phylogenetic approaches. Although Swadesh had proposed a method to estimate the divergence time of two languages, based on a molecular-clock assumption of lexical change (Swadesh 1950, 1952, 1955), he never expanded the model to account for more than two languages and preferred to draw maps of geographic and genetic proximity between multiple languages instead of forcing his data into a tree model (Swadesh 1959).

Dyen proposed a method by which a family tree could be inferred from data coded in Swadesh's lexicostatistical coding scheme (see Dyen in Hymes 1960), and Sankoff (Sankoff 1969) presented automatically computed trees with branch lengths using an agglomerative clustering procedure similar to UPGMA (Sokal and Michener 1958). Neither of the approaches, however, was extensively discussed or followed up in later work, and it was not until the end of the 20th century, when language trees gained new popularity due to the introduction of biological methods (Gray and Jordan 2000), that scholars began to reconstruct family trees *en masse* from linguistic data.

The introduction of phylogenetic methods from evolutionary biology to historical linguistics in the early 21st century marked a quantitative turn in the discipline (List 2014, p. 209f). Animated by the prospective of shedding light on long-standing so far unsolved problems in historical linguistics, more and more scholars began to use methods inspired or transferred from evolutionary biology to infer language trees (Ringe et al. 2002; Holman et al. 2011) or phylogenetic networks (Ben Hamed 2005; Nakhleh et al. 2005; Heggarty et al. 2010; Nelson-Sathi et al. 2011). A at the same time scholars began to produce new automatic methods for various tasks in historical linguistics, such as the comparison of words through phonetic alignments (Kondrak 2000; Prokić et al. 2009; List 2012a), the detection of homologous words (Steiner et al. 2011; List 2012b, 2014), or the inference of potential pathways of semantic change (Steiner et al. 2011; List et al. 2013).

While some scholars initially argued that the long-standing debate about waves and trees could be solved by reconstructing splits graphs, specifically Neighbor-Nets, from data on language distances (McMahon and McMahon 2005; Ben Hamed and Wang 2006; Heggarty et al. 2010), more and more scholars have by now turned to the reconstruction of language phylogenies based on likelihood models (Felsenstein 1981) and Bayesian inference using popular software packages such as MrBayes (Ronquist et al. 2009) and BEAST 2 (Bouckaert et al. 2014) which do not compute a single tree or network but rather provide a distribution (or forest) of trees which fit the data best under a certain model.

Although tree thinking is gaining ground in historical linguistics, there is still a considerable debate about the advantages of trees over distance-based approaches in historical linguistics (Jacques and List 2019; Kalyan and François 2019) and phylogenetic methods are still met with great reservation (Donohue et al. 2012; Hock 2017). In addition, scholars have pointed to problems in the data collection (Geisler and List 2010; Holm 2017), the phylogenetic models applied in phylogenetic reconstruction (Chang et al. 2015), or in the process of coding words for homology (List 2016; Wu and List 2021). While most of these problem can be handled by enhancing phylogenetic data, the way in which data are coded, and the models which are used to analyze the data in phylogenetic reconstruction, quite a few challenges remain, which will be discussed in detail in § 4.

6.3. Key Aspects of Language Change

While the discussion and inference of evolutionary patterns is considerably old in historical linguistics, the application for quantitative methods with the goal of aiding phylogetic reconstruction is still considerably young. Moreover, unlike in biology, where methods were designed for models which were originally designed for the handling of evolutionary processes in biology alone, most phylogenetic methods in linguistics have not created for linguistic purposes but were transferred from neighboring disciplines. While interdisciplinary transfer has many advantages, it also bears the risk of creating pointless results which are difficult to explain. In most cases, a careful modeling accompanied by an exhaustive review of the constructs proposed in the source discipline seems indiscpensible. Having provided a short overview on major historical aspects of evolutionary thinking in historical linguistics in the previous section, it is time to discuss some core aspects of language change which are important for a better understanding of language evolution in contrast to biological evolution: the systematicity of language change and specifically sound change (§ 3.1), the fact that words can have a history of their own (§ 3.2), the impact of communication strategies on language change processes (§ 3.3), and the possibility that human language is not monophyletic (§ 3.4).

6.3.1. It's the System!

Language is a system which essentially concatenates a fixed number of sounds to sequences, being only restricted by the encoding and decoding capacities of its users. This is the core reason why sound change is so different from change in biological characters. If we say that German d goes back to Proto-Germanic *θ (pronounced as *th* in *path*), this does not mean that there were a couple of mutations in a couple of words of the German language. Instead it means that the system which produced the words for Proto-Germanic changed the way in which the sound *θ was produced in the original system.

In some sense, we can think metaphorically of a typewriter, in which we replace a letter by another one. As a result, whenever we want to type a given word in the way we know it, we will type it with the new letter instead. But this analogy would be to restricted, as we can also add new letters to the typewriter, or remove existing ones. We can also split one letter key into two, as happens in the case of palatalization, which is a very common type of sound change during which sounds like [k] or [g] turn into sounds like [tʃ] and [dʒ] when being followed by front vowels (compare Italian *cento* “hundred,” which was pronounced [kɛntum] in Latin and is now pronounced as [tʃɛnto]).

Since it is the sound system that changes during the process we call sound change, and not the words (which are just a reflection of the output of the system), we cannot equate sound change with mutations in biological sequences, since mutations do not recur across all sequences in a genome, replacing one DNA segment by another one, which may not even have existed before. The change in the system, as opposed to the sequences that the system produces, is the reason for the apparent regularity of sound change.

This culminates in Leonard Bloomfield’s (1887-1949) famous expression that ‘phonemes [i. e., the minimal distinctive units of language] change’ (Bloomfield 1973, p. 351). From the perspective of formal approaches to sequence comparison, we could restate this as: ‘alphabets change.’ Hruschka et al. (2015) have compared sound change with concerted evolution in biology. We can state the analogy in simpler terms: sound change reflects systemic changes in language history, and concerted evolution results from systemic changes in biological evolution. It’s the system, stupid!

Given that sound systems change in language history, this means that the problem of character alignments (i.e. determining homology/cognacy) in linguistics cannot be directly solved with the same techniques that are used in biology, where the alphabets are assumed to be constant, and alignments are supposed to identify mutations alone. If we want to compare sequences in linguistics, where we have to compare sequences that were basically drawn from different alphabets, this means that we need to find out which sounds correspond to which sounds across different languages while at the same time trying to align them.

In Figure 3, I have tried to give a fictitious example for sound change that could occur with two speakers of the same language and would force their pronunciations to diverge. The first speaker suffers from a cold, which is why the person pronounces all nasal sounds [m, n, ŋ] (for convenience written as *m, n, ng*) as their corresponding voiced stop sounds [b, d, g]. The second speaker suffers from the loss of two front teeth, which is why this person pronounces instances of the sharp [s] as [θ] (for convenience written as *th*), while all voiced instances of the sound ([z]) become pronounced as a [ð] (for convenience written as *dh*). As can be seen from the figure, the actual comparison of the five test words which they procude cannot be done by relying on the similarity of individual segments, but needs to take the overal structure of them into account.

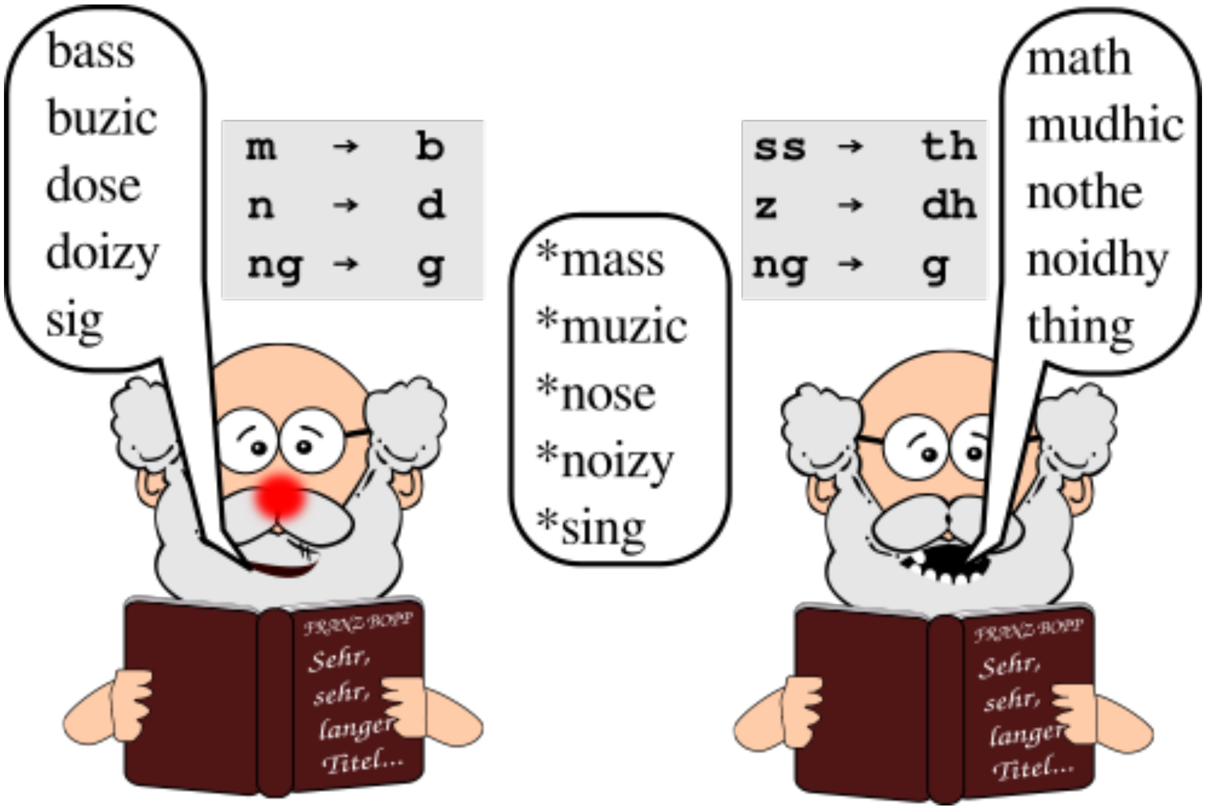


***Figure 3**: Fictitious example for the systemicity of sound change.*

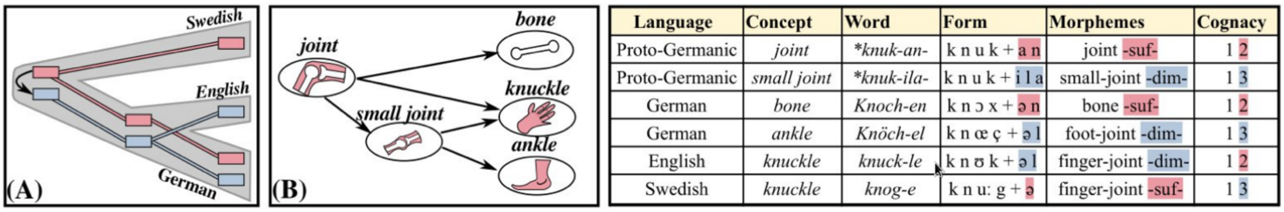
6.3.2. Chaque mot a son histoire

The competing models of the family tree and the wave theory determined a large part of the discussions about language evolution in the 19th century. At the same time, however, dialectologists had begun to investigate to which degree the *sound laws* which had been for the major Indo-European language family would be reflected in the contemporary dialects of European languages. Realizing that the supposed regularities would never match up completely and that exceptions could be found in all parts of the lexicon, they concluded that – at least to some degree – language evolution should not exclusively be studied as the evolution of entire languages, but rather as the individual evolution of the entities which define them.

Under the slogan “Chaque mot a son histoire” (*every word has its history*), usually attributed to Jules Gillieron (1854-1926, Campbell 1999, p. 189), they began to assemble evidence in the form of dialect maps in which the geographic distribution of words and sound correspondence patterns was displayed. Contrary to the promise of the slogan, however, concrete word histories involving concrete evolutionary scenarios of word evolution in time were rarely reconstructed.

Along with the quantitative turn, we can see an increased interest in the reconstruction of evolutionary scenarios for individual words (Gray et al. 2007; Gąsiorowski 2013) inspired by attempts to reconcile gene trees with species trees in evolutionary biology (Nakhleh 2013). Concrete attempts of automating the reconstruction of evolutionary scenarios at the word level, however, have failed to yield satisfying results so far, since they usually naively derive word trees from automatically inferred phonetic alignments and thus fail to distinguish processes of sound change from processes of morphological change (Willems et al. 2016).

Similar to genes in biology, which may evolve by random mutation and by more disruptive processes like duplication, the words of a language can evolve due to sound change and due to various processes of *word formation* by which existing language material is reused to form new words (List 2016). As a result of word formation, the words in a given language can be grouped into *word families* which share homologous material. As an example, consider German *Knochen* “bone” and Swedish *knoge* “knuckle,” which both back to Proto-Germanic *knuk-an “joint.” While the Proto-Germanic word is not reflected in English, German has another word *Knöchel* “knuckle,” which has a direct counterpart in English *knuckle* and goes back to Proto-Germanic *knuk-ila “small joint,” a diminutive form. Figure 3 shows how the complex etymological relations can be displayed in the form word trees which have been reconciled with the overall language tree.



Language	Concept	Word	Form	Morphemes	Cognacy
Proto-Germanic	*joint*	**knuk-an-*	k n u k + a n	joint -suf-	1 2
Proto-Germanic	*small joint*	**knuk-ila-*	k n u k + i l a	small-joint -dim-	1 3
German	*bone*	*Knoch-en*	k n ɔ x + ə n	bone -suf-	1 2
German	*ankle*	*Knöch-el*	k n œ ç + ə l	foot-joint -dim-	1 3
English	*knuckle*	*knuck-le*	k n ʊ k + ə l	finger-joint -dim-	1 2
Swedish	*knuckle*	*knog-e*	k n u: g + ə	finger-joint -suf-	1 3

Figure 4*: Word family for *knuk- “bone” in Proto-Germanic and its evolution in German, English, and Swedish. A shows the word tree which was reconciled with the language tree. B shows the semantic change pattern. The table shows the etymological relations among the words and their parts (morphemes).*

The example of “bone” in the Germanic languages show that in order to reconstruct evolutionary scenarios for individual words, a thorough distinction of sound change and word formation processes is needed. This has direct consequences on phonetic alignment analyses in linguistics, which can only model sound change processes and should therefore not be used in those cases where word formation processes obscure word similarities (Schweikhard and List 2020). As a result, however, the reconstruction of evolutionary scenarios is very tedious and although initial attempts to formalize the representation of etymological scenarios have been carried out (Schweikhard and List forthcoming), the future will show to which degree this procedure can be automated.

6.3.3. Speakers and Listeners

In contrast to evolution in biology, where organisms are reproduced physically, language exists only in communication, where it is constantly reproduced by those who produce speech and those who perceive speech. This speaker-listener or producer-recipient model (to include sign languages as one form of natural language which is neither produced by speaking nor perceived by listening) of language use is important for a closer understanding of the evolutionary dynamics underlying language change. Both perspectives, the perspective of the producer and the perspective of the recipient, invoke different dynamics that guide communication. From the perspective of the producer, it would be useful to reduce the efforts needed to transport a message to the recipient. As a result, speakers tend to assimilate words, shorten forms, or produce words imperfectly. Since the message needs to be decoded by the recipients, however, speakers cannot allow themselves to be too sloppy when producing speech signals. As a result, communication always oscillates between the two extremes of effortless speech on the one hand and clear messages coded in speech on the other hand.

The impact of these opposing forces on sound change are still not completely understood, but Ohala (1989) assumed that some concrete sound change phenomena might be better explained as originating from a misperception by speech recipients (pp. 182) rather than resulting from speech production, which is often assumed to be the most important factor contributing to sound change (Hock 1991, p. 11). Sound change induced by reception would result from acoustic similarities in spoken and from visual similarities in signed languages. On the other hand, sound change induced by production typically reflects similarities in articulation (of sounds or signs).

Unfortunately, no large-scale studies on sound change patterns in the languages of the world have been carried out so far. As a result, it remains an open question which of the two forces is prevailent in sound change. As a result, it is not possible to verify whether it is justified to identify production-driven changes with regular (systemic) language change (which happens gradually during acquisition, or – depending on the theory – also in later stages, Bybee 2002), and perception-based changes with change resulting from second language acquisition and language contact (Mowrey and Pagliuca 1995, p. 48).

6.3.4. Polygenesis is Possible

Biologists are often surprised when they hear that linguists are reluctant to apply methods for automatic language comparison to large and global language samples. In linguistics, scholars tend to emphasize that one can only compare languages historically (i.e., with the goal of investigating their evolutionary history) when knowing that they are *genetically related*. One could, of course, compare all languages in the world (and there are studies in which this is done, Jäger 2018), but a successful classical historical comparison of languages requires to know that they share a common origin. This knowledge can be achieved in different ways, and scholars in linguistics often differ in the details Campbell and Poser (2008), but they agree that the comparison of two random languages like, for example, Chinese and Greek (as it was carried out by Webb 1787) would not make sense if the goal consists in the identification of shared etymologies, since it is commonly assumed that the languages are not related with each other.

That linguists have to prove that two or more languages are related before they start to compare them is in strong contrast to biology. The idea of multiple origins as an alternative to a single origin itself has also been discussed in evolutionary biology. In linguistics, however, scholars are largely agnostic regarding the common origin of all languages, and even those who assume that languge originated only once usually acknowledge that the methods for historical language comparison are far too shallow to be able to reconstruct back to Proto-World.

Affirming disinterest in the origin of language has a long tradition in linguistics. In its *Statuts* from 1866 (published in 1871, 1871), the Société de Linguistique de Paris declared that it would not support any research on the origin of language. Even August Schleicher, who propagated language trees, affirmed this attitude in a letter to Ernst Haeckel (Schleicher 1863, p. 22), where he wrote:

> It is impossible to presuppose a material descent of all languages from a single proto-language. (My translation, original text: “Eine so zu sagen materielle

> Abstammung aller Sprachen von einer einzigen Ursprache können wir also unmöglich voraussetzen.”)

Although not explicitly spelled out nowadays, the *statutes* are still active in most linguistic circles. Even scholars who only try to bend the boundaries of the classical methods for historical language comparison and propose deeper groupings of well-established language families, like Indo-European and Basque (Blevins and Sproat 2021) or Turkic, Mongolic, Tungusic, Japanese, and Korean (known as “Altaic” language family) receive harsh and at times also justified criticism (Georg 2017).

6.4. Future Challenges for Historical Linguistics

Historical linguistics has made a lot of progress since the beginning of the quantitative turn. On the one hand, phylogenetic reconstruction has improved a lot and scholars begin to acknowledge the advantage of trees with branch lengths and methods that overcome the often circular reasoning of cladistic approaches to subgrouping. On the other hand, methods for automatic sequence comparison that can be used for phonetic alignment and the automatic detection of homologous words have not only greatly improved over the past years, but also been complemented by annotation tools which make the general process data annotation in historical linguistics much more promising (List 2017) and allow scholars to inspect their data in ways which would not have been possible without the help of interactive computational applications.

There remains, however, a considerable number of major future challenges for historical linguistics, which shall be discussed in a bit more detail in the following. These challenges pertain to the area of *linguistic typology*, a branch in comparative linguistics, which deals with the world-wide distribution of certain linguistic properties (so-called *universals*), or – in its dynamic form – with worldwide tendencies and trends of language change.

6.4.1. Typology of Sound Change

The first of these challenges is the challenge to establish a world-wide typology of sound change patterns. We have seen that sound change is a process that depends highly on the overall system of the language in which it occurs. Nevertheless, scholars have been noting for a long time now that sound changes proceed in surprisingly similar ways across independent language families spoken in different areas of the world. Initial attempts to establish a typology of sound change have been made in the past, but studies have been limited to certain types of sounds and areas in which the sounds occur Yang and Xu (2019). Another problem is that none of the approaches carried out so far attempted to unify transcription systems or to list sound changes exhaustively for a uniform sample of the languages in question.

As three major obstacles in the compilation of a typology of sound change, we can identify:

the *data* problem (it is tedious if not impossible to find a broad enough sample of sound change patterns for the more than 300 language families in the world,

the *modeling* (current formats lack the detail which we would need to handle sound change processes in a reasonable complexity, and

the *comparison* problem (it is not clear how to identify similar sound change processes in cross-linguistic samples, given the important role which systemic aspects play in triggering or blocking sound change).

Linguistic research has advanced with respect to some of these obstacles. Unified standards for lexical data have been proposed (Forkel et al. 2018) and have lead to the publication of large cross-linguistic lexical datasets (Rzymski et al. 2020). At the same time, standards for the representation of speech sounds have been proposed (Anderson et al. 2018) and successfully integrated into workflows for the retro-standardization of linguistic data (see Geisler et al. forthcoming for a recent example). Methods for automatic language comparison have also greatly improved lately, as demonstrated by recent experiments in which the regularity of sound change was used to predict words which had not yet been elicited in linguistic fieldwork with the help of the information from homologous words in a computer-assisted workflow with high accuracy (Bodt and List 2021). Unfortunately, however, not many datasets provide information on attested or reconstructed ancestral languages which would be important to test the direction of sound change processes.

At the same time, the problem of comparing sound change processes across different language families has barely been investigated. Since similar sound change patterns (a [t] might become a [d] in many different languages) often involve different conditioning contexts (the sound change could only happen between vowels or only at the beginning or the end of a word), there is no straightforward way to compare sound changes directly. In the future, more innovations regarding the representation and inference of sound change processes will be needed in order to advance our research on the typology of sound change.

6.4.2. Typology of Lexical Borrowing

Lexical borrowing, the process by which words are transferred from one language to another, is one of the most characteristic aspects of language change, and one of the reasons whey many words have a history of their own. For a long time, scholars have been speculating that there are certain limits to borrowing depending on the meaning which borrowed words express. Thus, while technical terms are easily borrowed, since they often enter a given language together with the technical innovation they refer to, words denoting body parts are rarely exchanged, since all languages tend have words that denote body parts, so borrowing would only occur in those situations where a larger proportion of the language users is multilingual and actively switches from one language to the other in their daily conversation.

Knowing that words denoting certain concepts are much less or much more frequently borrowed in the languages in the world would be very useful for historical language comparison, since it would help us to distinguish borrowed from inherited traits in those cases where evidence is sparse. While there have been intensive debates regarding the question if globally ranked lists for the borrowability of concepts could be established (Starostin 2007; Lee and Sagart 2008; Pozdniakov 2014; Zenner et al. 2014), no true large-scale studies on lexical borrowing have been carried out so far (with 40 languages only, the World Loanword Database by Haspelmath and Tadmor 2009 is unfortunately not large enough to provide sufficient evidence).

Methods for automatic borrowing detection have made some progress recently (Miller et al. 2020; List and Forkel 2021; see the overview in List 2019). All in all, however, they are not yet capable of handling the complexities of lexical borrowing processes. Major improvements for automatic borrowing detection methods in the future can only be expected if scholars find ways to to infer the *direction* of borrowings and if they manage to design methods that could also detect instances of *semantic borrowing* or *loan transfer* in which words are translated from another language rather than being transferred directly (compare German *Wolkenkratzer*, lit. “cloud-scraper” → “skyscraper”).

6.4.3. Typology of Semantic Shift

Words do not only modify their sounds during language evolution, they may also easily change their meanings. This process, known as *semantic change* can lead to interesting extreme cases of semantic divergence, as in the case of German *selig* which is homologous with English *silly*, but means “holy” rather than “crazy” (Szemerényi 1970). The idiosyncrazy of the processes by which words change their meaning are one of the factors that contribute to the difficulty of finding homologous words across different languages. The ultimate source of semantic change lies in the actual communication involving language producers and language recipients (as discussed in § 3.3) during which creative language used is conventionalized with time (Enfield and Comrie 2015). Scholars have noted for a long time that the pathways of semantic change may resemble across languages used in different areas in the world and stemming from different language families. Initial attempts to collect common pathways of semantic change across a larger sample of the world’s languages exist, but they are usually limited in scope (Wilkins 1996; Blank 1997). Even the *Database of Semantic Shifts* by Zalizniak et al. (2020), which assembles semantic shift patterns for more than 2000 concepts from the literature is still rather small in size and extremely sparse, since the authors rely on direct quotes from the literature.

Similar to the problem of establishing a sound change typology, we can identify three major problems in establishing a large, world-wide sample of semantic shifts:

the *data* problem (homogeneous, standardized lists of lexical items for a large sample of the world’s languages would be needed)

the *modeling* problem (concepts would have to be modeled rigorously and unified across all data)

the *inference* problem (there are not enough ancestral languages to infer a sufficiently large amount of semantic shift patterns from known processes of semantic change)

Some progress has been made by now. The data problem has been partially already addressed with the large standardized collections mentioned above (Rzymski et al. 2020), which can also be used to infer polysemy networks from cross-linguistic data (List et al. 2018) which are undirected, but seem to provide some hints on pathways of semantic change (Münch and Dellert 2015). The modeling problem has also found a preliminary solution in the form of the Concepticon project (List et al. 2021), a database which offers standard concept identifiers along with definitions and links to existing datasets in which the concepts have been used (https://concepticon.clld.org). For the inference problem, Dellert and Buch (2016) has proposed a method to infer the direction of semantic change processes by applying causal inference techniques, but it seems that the approach is very data hungry, so it would not be feasible to employ it with sparse cross-linguistic collections of lexical data, as they have been assembled by now.

There remain, however, many challenges for future work. Cross-linguistic data needs to be expanded, the representation of concepts needs to be expanded as well in order to provide a broader coverage, and inference methods by which semantic change directions can be inferred from cross-linguistic data for individual language families need to be developed. Here, it might be possible to make use of methods for ancestral state reconstruction (Jäger and List 2018) applied to automatically inferred sets of homologous words, but this would also presuppose substantially enhanced algorithms that infer homologous words which express different meanings (*cross-semantic cognates*). While initial methods have been proposed for this task (Wu et al. 2020), they have so far not been rigorously evaluated.

6.5. Conclusion

Although the processes driving language evolution are not the same as those driving biological evolution, both kinds of evolution have a lot in common, which is also reflected in the successful transfer of methods for phylogenetic reconstruction in biology to the domain of language. While trees and alternative models of macro-evolution were discussed quite early in the history of linguistics, it was not until the end of the 20th century that scholars increased their efforts to formalize their approaches. Interdisciplinary exchange with evolutionary biology and transfer of methods from biology to linguistics played a crucial role in the propagation of quantitative methods in historical linguistics.

Despite numerous commonalities, however, biological evolution and linguistic evolution also exhibit striking differences. Four of these differences – the systemic nature of sound change (§ 3.1), the individual histories of words (§ 3.2), the interaction of speaker and listener (§ 3.3), and the possibility that spoken languages have evolved multiple times (§ 3.4) – were discussed more closely in this study.

With these differences in mind, I have then presented three major challenges for the field of historical linguistics, for which no solutions have been proposed so far, and for which it is also unlikely that we will find a satisfying solution during the next time. These problems – the identification of global tendencies in sound change (§ 4.1), the ranking of concepts according to their borrowability (§ 4.2), and the assessments of general tendencies in semantic change (§ 4.3) are specifically challenging because they result at least in part from the peculiarities of language evolution. As a result, it is not possible to solve these problems by simply transferring methods from evolutionary fields. Instead, targeted solutions will be needed which combine current efforts to automate the traditional linguistic methods for historical language comparison with new ideas derived from interdisciplinary exchange with different disciplines. While the transfer of biological methods has played an important role for linguistic approaches in the past, it seems that it is time to work towards a more conscious adaptation of methods to our linguistic needs.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 715618), CALC, *Computer-Assisted Language Comparison* (https://digling.org/calc).

References

Anderson C, Tresoldi T, Chacon TC, et al (2018) A Cross-Linguistic Database of Phonetic Transcription Systems. Yearbook of the Poznań Linguistic Meeting 4:21–53

Ben Hamed M (2005) Neighbour-Nets portray the Chinese dialect continuum and the linguistic legacy of Chinaś demic history. Proceedings of the Royal Society of London B: Biological Sciences 272:1015–1022

Ben Hamed M, Wang F (2006) Stuck in the forest: Trees, networks and Chinese dialects. Diachronica 23:29–60

Blank A (1997) Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen. Niemeyer, Tübingen

Blevins J, Sproat R (2021) Statistical evidence for the proto-indo- european-euskarian hypothesis a word-list approach integrating phonotactics. Diachronica 0:1–59. https://doi.org/10.1075/dia.19014.ble

Bloomfield L (1973) Language. Allen & Unwin, London

Bodt TA, List J-M (2021) Reflex prediction. A case study of western kho-bwa. Diachronica 0:1– 38

Bouckaert R, Heled J, Kühnert D, et al (2014) BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Computational Biology 10:e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Bouckaert R, Lemey P, Dunn M, et al (2012) Mapping the origins and expansion of the Indo-European language family. Science 337:957–960

Bouckaert RR, Bowern C, Atkinson QD (2018) The origin and expansion of pama–nyunganlanguages across australia. Nature Ecology and Evolution 2:741–749. https://doi.org/https://doi.org/10.1038/s41559-018-0489-3

Bybee JL (2002) Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. Language Variation and Change 14:261–290

Campbell L (1999) Historical linguistics. An introduction, 2nd edn. Edinburgh Univ. Press, Edinburgh

Campbell L, Poser WJ (2008) Language classification: History and method. Cambridge University Press, Cambridge

Chang W, Cathcart C, Hall D, Garret A (2015) Ancestry-constrained phylogenetic analysis ssupport the Indo-European steppe hypothesis. Language 91:194–244

Dellert J, Buch A (2016) Using computational criteria to extract large Swadesh Lists for lexicostatistics. In: Bentz C, Jäger G, Yanovich I (eds) Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics. Eberhard-Karls University, Tübingen

Donohue M, Denham T, Oppenheimer S (2012) New methodologies for historical linguistics? Calibrating a lexicon-based methodology for diffusion vs. subgrouping. Diachronica 29:505–522

Dybo A, Starostin GS (2008) In defense of the comparative method, or the end of the Vovin controversy. In: Smirnov IS (ed) Aspekty komparativistiki. RGGU, Moscow, pp 119–258

Enfield NJ, Comrie B (eds) (2015) Languages of Mainland South-East Asia. The state of the art. Mouton de Gruyter, Berlin; New York

Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. Journal of Molecular Evolution 17:368–376

Forkel R, List J-M, Greenhill SJ, et al (2018) Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. Scientific Data 5:1–10

Fox A (1995) Linguistic reconstruction: An introduction to theory and method. Oxford University Press, Oxford

Gąsiorowski P (2013) Gruit grus: The indo-european names of the crane. Studia Etymologica Cracoviensia 18:51–68. https://doi.org/10.4467/20843836SE.13.003.0940

Geisler H, Forkel R, List J-M (forthcoming) A digital, retro-standardized edition of the tableaux phonétiques des patois suisses romands (TPPSR). In: Avanzi M, LoVecchio N, Millour A, Thibault A (eds) Nouveaux regards sur la variation dialectale. Éditions de Linguistique et de Philologie, Strasbourg, pp 1–21

Geisler H, List J-M (2013) Do languages grow on trees? The tree metaphor in the history of linguistics. In: Fangerau H, Geisler H, Halling T, Martin W (eds) Classification and evolution in biology, linguistics and the history of science. Concepts – methods – visualization. Franz Steiner Verlag, Stuttgart, pp 111–124

Geisler H, List J-M (2010) Beautiful trees on unstable ground. Notes on the data problem in lexicostatistics. In: Hettrich H (ed) Die Ausbreitung des Indogermanischen. Thesen aus Sprachwissenschaft, Archäologie und Genetik. Reichert, Wiesbaden

Georg S (2017) The role of paradigmatic morphology in historical, areal and genealogical linguistics. Journal of Language Contact 10:353–381

Gerardi F, Reichert S, Coelho Aragon C (2021) TuLeD (tupían lexical database): Introducing a database of a south american language family. Language Resources and Evaluation 55:1–19. https://doi.org/10.1007/s10579-020-09521-5

Gray RD, Drummond AJ, Greenhill SJ (2009) Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. Science 323:479–483

Gray RD, Greenhill SJ, Ross RM (2007) The pleasures and perils of Darwinzing culture (with phylogenies). Biological Theory 2:360–375

Gray RD, Jordan FM (2000) Language trees support the express-train sequences of Austronesian expansion. Nature 1052–1055

Grimm J (1822) Deutsche Grammatik, 2nd edn. Dieterichsche Buchhandlung, Göttingen Haspelmath M, Tadmor U (eds) (2009). de Gruyter, Berlin; New York

Heggarty P, Maguire W, McMahon A (2010) Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories. Philos Trans R Soc Lond, B, Biol Sci 365:3829–3843

Hock HH (2017) Indo-European linguistics meets Micronesian and Sunda-Sulawesi. Wellington Working Papers in Linguistics 23:63–67

Hock HH (1991) Principles of historical linguistics, 2nd edn. Mouton de Gruyter, Berlin

Holm HJ (2017) Steppe Homeland of Indo-Europeans Favored by a Bayesian Approach with Revised Data and Processing. Glottometrics 37:54–81

Holman EricW, Brown CH, Wichmann S, et al (2011) Automated dating of the worldś language families based on lexical similarity. Current Anthropology 52:841–875

Hrozný B (1915) Die Lösung des hethitischen Problems. Mitteilungen der Deutschen Orient-Gesellschaft 56:17–50

Hruschka DJ, Branford S, Smith ED, et al (2015) Detecting regular sound changes in linguistics as events of concerted evolution. Curr Biol 25:1–9

Hübschmann H (1877) Ueber die stellung des armenischen im kreise der indogermanischen sprachen. Zeitschrift für vergleichende Sprachforschung der indogermanischen Sprachen 23:5– 49

Hymes DH (1960) Comment on "Lexicostatistics so far" (by D. H. Hymes). Current Anthropology 1:34–39

Jacques G, List J-M (2019) Save the trees: Why we need tree models in linguistic reconstruction (and when we should apply them). Journal of Historical Linguistics 9:128–166. https://doi.org/https://doi.org/10.1075/jhl.17008.mat

Jäger G (2018) Global-scale phylogenetic linguistic inference from lexical resources. Scientific Data 5:1–16

Jäger G, List J-M (2018) Using ancestral state reconstruction methods for onomasiological reconstruction in multilingual word lists. Language Dynamics and Change 8:22–54

Kalyan S, François A (2019) When the waves meet the trees. Journal of Historical Linguistics 9:167–176. https://doi.org/10.1075/jhl.18019.kal

Klein WP (2004) Was wurde aus den Wörtern der hebräischen Ursprache? Zur entstehung der komparativen linguistik aus dem geist etymologischer spekulation. In: Veltri G, Necker G (eds) Gottes Sprache in der philologischen Werkstatt: Hebraistik vom 15. Bis zum 19. jahrhundert. procsym; Brill, Leiden, pp 3–23

Kolipakam V, Jordan FM, Dunn M, et al (2018) A Bayesian phylogenetic study of the Dravidian language family. Royal Society Open Science 5:1–17

Kondrak G (2000) A new algorithm for the alignment of phonetic sequences. In: Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference. pp 288–295

Kümmel MJ (2008) Konsonantenwandel. Reichert, Wiesbaden

Lee Y-J, Sagart L (2008) No limits to borrowing: The case of Bai and Chinese. Diachronica 25:357–385

List J-M (2012a) Multiple sequence alignment in historical linguistics: A sound class based approach. In: Boone E, Linke K, Schulpen M (eds) Proceedings of ConSOLE XIX. pp 241–260

List J-M (2019) Automated methods for the investigation of language contact situations, with a focus on lexical borrowing. Language and Linguistics Compass 13:1–16

List J-M (2017) A web-based interactive tool for creating, inspecting, editing, and publishing etymological datasets. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. System Demonstrations. Association for Computational Linguistics, Valencia, pp 9–12

List J-M (2016) Beyond cognacy: Historical relations between words and their implication for phylogenetic reconstruction. Journal of Language Evolution 1:119–136. https://doi.org/10.1093/jole/lzw006.

List J-M (2014) Sequence comparison in historical linguistics. Düsseldorf University Press, Düsseldorf

List J-M (2012b) LexStat. Automatic detection of cognates in multilingual wordlists. In: Proceedings of the EACL 2012 Joint Workshop of Visualization of Linguistic Patterns and Uncovering Language History from Multilingual Resources. Stroudsburg, pp 117–125

List J-M, Forkel R (2021) Automated identification of borrowings in multilingual wordlists [version 1; peer review: Awaiting peer review]. Open Research Europe 1:1–11

List J-M, Greenhill SJ, Anderson C, et al (2018) CLICS2. An improved database of cross-linguistic colexifications assembling lexical data with help of cross-linguistic data formats. Linguistic Typology 22:277–306

List J-M, Pathmanathan JS, Lopez P, Bapteste E (2016) Unity and disunity in evolutionary sciences: Process-based analogies open common research avenues for biology and linguistics. Biology Direct 11:1–17

List JM, Rzymski C, Greenhill S, et al (2021) Concepticon. A resource for the linking of concept lists. Version 2.5.0. Version 2.5.0. Max Planck Institute for the Science of Human History, Jena. URL https://concepticon.clld.org/

List J-M, Terhalle A, Urban M (2013) Using network approaches to enhance the analysis of cross-linguistic polysemies. In: Proceedings of the 10th International Conference on Computational Semantics – Short Papers. Association for Computational Linguistics, Stroudsburg, pp 347–353

McMahon A, McMahon R (2005) Language classification by numbers. Oxford University Press, Oxford

Miller JE, Tresoldi T, Zariquiey R, et al (2020) Using lexical language models to detect borrowings in monolingual wordlists. PLOS One 15:e0242709

Mowrey R, Pagliuca W (1995) The reductive character of articulatory evolution. Rivista di Linguistica 7:37–124

Münch A, Dellert J (2015) Evaluating the potential of a large-scale polysemy network as a model of plausible semantic shifts. In: Baayen H, Jäger G, Köllner M, et al. (eds) Proceedings of the 6th Conference on Quantitative Investigations in Theoretical Linguistics. Eberhard-Karls University, Tübingen

Nakhleh L (2013) Computational approaches to species phylogeny inference and gene tree reconciliation. Trends in Ecology and Evolution 28:719–728

Nakhleh L, Ringe D, Warnow T (2005) Perfect Phylogenetic Networks: A new methodology for reconstructing the evolutionary history of natural languages. Language 81:382–420

Nelson-Sathi S, List J-M, Geisler H, et al (2011) Networks uncover hidden lexical borrowing in Indo-European language evolution. Proceedings of the Royal Society of London B: Biological Sciences 278:1794–1803

Ohala JJ (1989) Sound change is drawn from a pool of synchronic variation. In: Breivik LE, Jahr EH (eds) Language Change: Contributions to the study of its causes. Mouton de Gruyter, Berlin, pp 173–198

Pozdniakov K (2014) O poroge rodstva i indekse stabil’nosti v bazisnoj leksike pri massovom sravnenii: Atlantičeskie jazyki. Journal of Language Relationship 11:187–237

Prokić J, Wieling M, Nerbonne J (2009) Multiple sequence alignments in linguistics. In: Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education. pp 18–25

Rask RK (1818) Undersögelse om det gamle Nordiske eller Islandske sprogs oprindelse. Gyldendalske Boghandlings Forlag, Copenhagen

Ringe D, Warnow T, Taylor A (2002) Indo-European and computational cladistics. Transactions of the Philological Society 100:59–129

Ronquist F, Mark P van der, Huelsenbeck JP (2009) Bayesian phylogenetic analysis using MrBayes. In: Lemey P, Salemi M, Vandamme A-M (eds) The phylogenetic handbook. A practical approach to phylogenetic analysis and hypothesis testing, Second Edition. Cambridge University Press, Cambridge, pp 210–266

Rzymski C, Tresoldi T, Greenhill S, et al (2020) The Database of Cross-Linguistic Colexifications, reproducible analysis of cross- linguistic polysemies. Scientific Data 7:1–12

Sagart L, Jacques G, Lai Y, et al (2019) Dated language phylogenies shed light on the ancestry of Sino-Tibetan. Proceedings of the National Academy of Science of the United States of America 116:10317–10322

Sankoff D (1969) Historical linguistics as stochastic process. Dissertation, McGill University

Schleicher A (1863) Die Darwinsche Theorie und die Sprachwissenschaft: Offenes sendschreiben an herrn dr. Ernst haeckel. Hermann Böhlau, Weimar

Schleicher A (1861) Compendium der vergleichenden Grammatik der indogermanischen Sprache. Böhlau, Weimar

Schleicher A (1853) O jazyku litevském, zvláště ohledem na slovanský. Čteno v posezení sekcí filologické král. České Společnosti Nauk dne 6. Června 1853. Časopis Českého Museum 27:320–334

Schmidt J (1872) Die Verwantschaftsverhältnisse der indogermanischen Sprachen. Hermann Böhlau, Weimar

Schuchardt H (1900) Über die Klassifikation der romanischen Mundarten. Probe-Vorlesung, gehalten zu Leipzig am 30. April 1870. Graz

Schweikhard NE, List J-M (forthcoming) Modeling word trees in historical linguistics. Preliminary ideas for the reconciliation of word trees and language trees. In: Brogyanyi B, Lipp R (eds) Sprach(en)forschung: Disziplinen und interdisziplinarität. Akten der 27. Fachtagung der gesellschaft für sprache und sprachen. Dr. Kovač, Hamburg, pp 1–20

Schweikhard NE, List J-M (2020) Developing an annotation framework for word formation processes in comparative linguistics. SKASE Journal of Theoretical Linguistics 17:2–26

Societé de Linguistique de Paris (1871) Statuts: Approuvés par décision ministérielle du 8 mars 1866. Bulletin de la Société de Linguistique de Paris 1:III–IV

Sokal RobertR, Michener CharlesD (1958) A statistical method for evaluating systematic relationships. University of Kansas Scientific Bulletin 28:1409–1438

Starostin SA (2007) Computer-based simulation of the glottochronological process (Letter to M. Gell-Mann). In: S. A. Starostin: Trudy po yazykoznaniyu [S. A. Starostin: Works in Linguistics]. LRC Publishing House, pp 854–861

Steiner L, Stadler PF, Cysouw M (2011) A pipeline for computational historical linguistics. Language Dynamics and Change 1:89–127

Swadesh M (1959) The Mesh Principle in Comparative Linguistics. Anthropological Linguistics 1:7–14

Swadesh M (1955) Towards greater accuracy in lexicostatistic dating. International Journal of American Linguistics 21:121–137

Swadesh M (1952) Lexico-statistic dating of prehistoric ethnic contacts: With special reference to north american indians and eskimos. Proceedings of the American Philosophical Society 96:452–463

Swadesh M (1950) Salish internal relationships. International Journal of American Linguistics 16:157–167

Szemerényi O (1970) Einführung in die vergleichende Sprachwissenschaft. Wissenschaftliche Buchgesellschaft, Darmstadt

Webb D (1787) Some reasons for thinking, that the Greek language was borrowed from the Chinese. In: Notes on the Grammatica Sinica of Mons. Fourmont. Dodsley, London

Wilkins DP (1996) Natural tendencies of semantic change and the search for cognates. In: Durie M (ed) The comparative method reviewed: Regularity and irregularity in language change. Oxford University Press, New York, pp 264–304

Willems M, Lord E, Laforest L, et al (2016) Using hybridization networks to retrace the evolution of Indo-European languages. BMC Evolutionary Biology 16:1–18. https://doi.org/10.1186/s12862-016-0745-6

Wu M-S, List J-M (2021) Annotating cognates in phylogenetic studies of south-east asian languages. Humanities Commons 0:1–31

Wu M-S, Schweikhard NE, Bodt TA, et al (2020) Computer-assisted language comparison. State of the art. Journal of Open Humanities Data 6:1–14

Yang C, Xu Y (2019) A review of tone change studies in East and Southeast Asia. Diachronica 36:417–459

Zalizniak AA, Smirnitskaya A, Russo M, et al (2020) Database of semantic shifts (version from 07/10/2020). Version 2020-10-07. Institute of Linguistics at the Russian Academy of Sciences, Moscow. URL http://datsemshift.ru/

Zenner E, Speelman D, Geeraerts D (2014) Core vocabulary, borrowability and entrenchment. Diachronica 31:74–105.

Chapter 7

A community science model for inter-disciplinary evolution education and school improvement

Dustin Eirdosh[i], Susan Hanisch[ii]

Abstract

A generalized conceptualization of evolutionary processes allows for a view of the cognitive, behavioral, and cultural variation in our everyday lives as elements of diverse evolving systems. Such a view invites questions about how cultural evolutionary processes may favor or hinder the expression of variant thoughts and behaviors, any of which may be more or less valued by any given community. From an educational perspective, this implies an untapped potential for engaging students in understanding the cultural evolutionary dynamics of their everyday lives, schools, and broader communities. As a strategy to engage this potential, the Community Science Lab at the Max Planck Institute for Evolutionary Anthropology is developing a unique model of *Community-Based Cultural Evolution* (CBCE) for inter-institutional collaboration at the intersection of evolution education and applied school improvement efforts. Using advances in teaching for conceptual understanding and transfer of learning, the CBCE model aims to empower students to clarify, investigate, and collaboratively influence the cultural evolutionary dynamics of their own school and surrounding communities. The relationship between students' evolving intuitive theories of school improvement, and the evolving scientific theories of school improvement scientists, provides a framework for understanding the development of student conceptions of cultural (and, perhaps, biological) change more generally. This chapter provides a conceptual foundation for exploring the claim that engaging students in reflecting on the cognitive, behavioral, and cultural evolutionary processes in their everyday lives provides new opportunities for school improvement and interdisciplinary evolution education initiatives. The practical and systemic challenges of this approach are clarified and future directions are outlined.

---

[i] Max Planck Institute for Evolutionary Anthropology, Department of Comparative Cultural Psychology, University of Jena, Biology Education Working Group.

[ii] Max Planck Institute for Evolutionary Anthropology, Department of Comparative Cultural Psychology, University of Leipzig, Faculty of Education

7.1. Introduction

A significant trend in 21st century educational innovation is to involve students and other stakeholders (teachers, parents, community members) in the participatory improvement of school design (Mintrop 2016). An obvious question in such initiatives would be: what skills do such stakeholders need in order to effectively engage as participants in school improvement aims? For many education leaders, the answer to this question may take the form of broad competencies such as critical thinking or social-emotional learning. Apparently missing from this current discussion, however, is consideration of the conceptual understandings that different stakeholders might have in regards to the nature of human learning (and human behavior, cognition, or culture more generally) within present-day societies. That is, current work in participatory school improvement overwhelmingly lacks frameworks or resources for deep engagement with stakeholders about big questions regarding the origins and purpose of schools in society and the relationship between schools and the human condition. In this chapter, we argue that generalizable evolutionary concepts may be practically helpful for the aims of supporting community-based approaches to school improvement, especially when embedded within an interdisciplinary evolution science curriculum.

First, we provide a brief review of the state of the art in thinking about evolutionary concepts as related to school improvement. Here we describe that while the field of evolution education has not significantly engaged with evolutionary perspectives on learning, teaching, or schooling, the field of educational psychology has developed a diversity of often disparate and possibly incongruent evolutionary conceptualizations for school improvement.

From this brief overview, we then highlight our aims and early work in establishing a community science model for applied educational design research in this space, which we call *Community-Based Cultural Evolution*. We then present the Evolving Schools project as an exploratory example for organizing school-based research collaborations. We conclude with challenges and future directions.

7.2. Evolutionary Concepts in School Improvement

Educational thinkers, from across cultures and historical periods in which schools have existed, have commonly thought about teaching and schooling in relation to some conceptualization of the human condition more broadly. With the emergence and popularization of Charles Darwin's work concurrently with the rising global trend of universal formal schooling, it is perhaps not surprising that even early influential educational theorists such as Maria Montessori (see Frierson 2018) and John Dewey (see Popp 2012) worked intensively to contextualize their thinking within the evolutionary science of their day. While education and evolution sciences have significantly diverged during the 20th and 21st centuries, there remains a complex, if highly fragmented, relationship between these fields. This complexity and fragmentation may also drive the diversity of views on how education and evolution could or should be related. The scope here is not to provide a comprehensive overview of this diversity, but only to characterize and briefly highlight some examples across the fields of evolution education and school improvement as a context for our more integrated approach.

7.2.1. The view from evolution education

It is uncontroversial to describe that the topic of *school improvement from an evolutionary perspective* is not on the radar for the international field of evolution education, which is a sub-field specialization of science and biology education. Specialized journals including *Evolution: Education & Outreach*, *Science & Education*, *Journal of Biological Education*, and *The American Biology Teacher* have limited relevant literature on record (the only even partially relevant examples we have been able to find include Gray 2011; Grobstein & Lesnick 2011; Eirdosh & Hanisch 2020). This is not surprising, as the field of evolution education overwhelmingly lacks a focus or research infrastructure for engaging general education students in the evolution of human behavior, cognition, and culture, more generally (Ziadie & Andrews 2019). It is beyond the scope here to explore the reasons why this may be the case, historically and conceptually (see Hanisch & Eirdosh 2020a for relevant discussions), only to point out that it is the current status quo. This is an important context, as our community science model includes an explicit emphasis on the conceptual content of evolution education as a driving support for student participation in school improvement.

7.2.2. Diverse traditions of evolutionary theorizing in school improvement

While evolution education may not engage with broader school improvement aims, educational psychologists and school improvement scientists have significantly engaged biological and generalized evolutionary theory as contexts for developing educational theories. Similarly to Nettle's (this volume) more general contextualization of evolutionary theory in psychological sciences, we have previously argued that all of educational theorizing is evolutionary in some sense (Eirdosh & Hanisch 2020b). The question, therefore, is not so much *if* a theory is evolutionary, but in what ways does it engage evolutionary concepts, and to what scientific and/or practical benefit?

Below, seven popularized theories or frameworks are concisely summarized, without analysis or judgement as to the scientific merits. The aim of this section is only to briefly highlight the diversity and largely fragmented state of evolutionary theorizing in educational research and school improvement literature. Overall, it should be noted that educational theory currently involves the application of generalized evolutionary concepts, especially within traditions of *Cognitive Load Theory*, *Prosocial Schools*, and *Networked Improvement Communities*.

Biologically primary and secondary learning. Evolutionary educational psychologist, David Geary (2005), has outlined an argument for distinguishing between domains of learning that humans have plausible (genetic) adaptations for acquiring through automatic, intuitive processes (e.g. speaking), and domains for which we likely do not have such evolved capacities (e.g. reading). Geary argues that learning of the former does not require instruction and can not be taught, while learning of the latter often does require or benefit from methods of direct instruction from experts.

Cognitive Load Theory. The career work of John Sweller (2004, 2008) and colleagues (Sweller & Sweller 2006) has yielded the influential and empirically supported Cognitive Load Theory for instructional design. Sweller builds on the work of David Geary's biologically primary versus secondary learning distinction, yet takes evolutionary theorizing in a more generalized direction as well. By conceptualizing the cognitive architecture of the (human) mind as a *Natural Information Processing System*, Sweller argues that a partial analogy can be drawn to processes between genetic and cognitive evolution, and that this analogy is central to understanding the constraints and requirements of effective instruction.

Self-Directed Education. Championed by evolutionary educational psychologist, Peter Gray (2011), Self-Directed Education employs an evolutionary theoretical narrative about the role of intrinsic human motivations to learn in mixed-age autonomy-supportive social environments to argue for a significant departure from the more rigid and hierarchical institutions of most modern schooling paradigms.

Self-Determination Theory. This theory of human motivation and psychological needs, developed by Deci & Ryan (2011), has been very influential in educational research towards the design of learning environments that maximise intrinsic motivation and well-being in students and teachers. Ryan & Hawley (2016) have contextualized the evolutionary origins and functions of the psychological needs of autonomy, competence, and relatedness in human motivation and well-being, as posited by Self-Determination Theory.

Prosocial Schools. Integrating generalized evolutionary theories with theories of collective action and psychological flexibility, Prosocial Schools is an organizational circle within Prosocial World (based on Atkins et al. 2019) which uses perspectives in cooperation science to synthesize educational innovations and network school improvement stakeholders. In this model, the eight *Core Design Principles* for cooperation originally identified by Elinor Ostrom and colleagues, and later generalized (Wilson et al. 2013), serve as a conceptual framework for the synthesis and design of participatory school improvement projects. We have previously discussed the complex relationship between the generalized conceptualization of evolution in Prosocial and the mainstream (gene-centric) conceptualization of evolution in general education (see Eirdosh & Hanisch 2020a).

Visible Learning. Educational researcher, John Hattie, has conducted the largest meta-analysis of educational efficacy research, organizing his insights into a synthesis he calls *Visible Learning*. Together with learning scientist, Gregory Yates (see Hattie & Yates 2013), Hattie has outlined a theoretical context for their synthesis of best practices, that is grounded in a socio-cognitive view of learning and evolution very congruent with, yet distinct from, the knowledge synthesis being advanced by Prosocial Schools.

Networked Improvement Communities. The influential school improvement strategy of *Networked Improvement Communities* (see Bryk et al. 2015) is broadly contextualized in evolutionary terms and has been developed on an explicitly generalized model of social learning as an evolutionary process, as conceptualized by learning theorist Douglas Engelbart (1962). In this model, collective human intelligence is seen as co-evolving with technology to support the kind of multilevel networked improvement communities that are at the heart of this school improvement model.

Each of these examples employs evolutionary concepts in sometimes similar, but often very different ways, conceptualized from different traditions across various disciplines. Many, if not all, of these theories *may* be argued to be congruent with each other along some dimensions, able to be integrated into a larger, more coherent generalized evolutionary theoretical framework. However, in practice, the theories above represent largely (though not entirely) disconnected education and research communities, with sometimes incongruent conclusions about educational design. For example, *Self-Directed Education* theorists and *Cognitive Load* theorists may have varyingly strong disagreements about the role of direct instruction and structured curricula in modern schooling. Prosocial Schools is one notable outlier here, in that this community explicitly seeks knowledge synthesis across many of these diverse traditions. Overall, we see this broader fragmentation and potential discordance across evolutionary theories of schooling as an opportunity for scientific learning and development.

7.2.3. Towards a more generalized, integrative, pluralistic, and participatory approach

The fractionated state of evolutionary theory within school improvement literature suggests a possible opportunity in exploring how a more systematic approach to *coherent theory building* (see Hanisch & Eirdosh, this volume) may offer value to diverse education stakeholders. We suggest that such an approach should include at least four core design commitments:

Generalized

In line with educational design research (McKenney & Reeves 2018), a community science model for school improvement should be developed across *multiple levels of abstraction*, from highly *generalized principles* that apply across widely diverse contexts, to supports for *local theorizing* about adaptation in local contexts. This aids the potential *compatibility* and *tolerance* of the model to be applied with integrity across diverse school communities.

Further, we suggest that a community science model should be built around critical reflection on the generalizability of core evolutionary concepts and conceptual relationships. That is, rather than embracing one particular tradition in the expansive landscape of evolutionary theorizing described above, a community-based model can engage students and school community members in reflective analysis of the scientific or practical value of generalizing evolutionary concepts in specific contexts.

Integrative

Given the expansive scope of relevance of evolutionary theorizing in school improvement, approaches should be integrative, seeking to empower educators and students to “work smarter, not harder”, and “do more by doing less”. That is, evolutionary approaches to school improvement offer potential that spans academic learning, social-emotional development, and sustainable community development. For that reason, emphasis should be placed on the potential for pursuing multiple aims within interdisciplinary programs, rather than continuing to develop fragmented programs across these areas. This can be done intentionally within the context of addressing the critical challenge of *curriculum overload* (OECD 2020) in school improvement.

Pluralistic

This expansive scope also yields the greatest challenge in interdisciplinary evolution education: the diversity of expert and novice perspectives on the nature and value of evolutionary theory (and theorizing in general) in the improvement of schools or society. That is, many individuals from many different backgrounds have many, often strong, opinions on what and how school improvement should look like, and if or how evolutionary concepts should be integrated in such efforts. For this reason, a community science model for school improvement must embrace a strategic pluralism (see Lohse, this volume). We specifically suggest a stronger emphasis on teaching about the *Nature of Concepts* as part of *Nature of Science* pedagogy. That is, the challenges of advancing a pluralism that is not relativistic (sensu Van Bouwel & Weber, 2008) requires deeper reflection on the nature of scientific explanations and the concepts we employ to construct such explanations. For that reason, engaging students in the critical analysis of the explanatory value of generalizing evolutionary concepts represents a novel pedagogical approach to navigating the conceptual diversity of the science.

Participatory

Finally, the novelty and complexity of advancing a community science model requires a deeply participatory approach to the design and spread of innovations (Atkins et al 2019; Mintrop 2018; Boyd 2014). Many schools embrace some degree or dimension of participation in school improvement, but this varies widely, and there remains little consensus or practical guidance regarding the conceptual learning dimensions of what students might need for, or gain from, engaging as participants in school improvement processes.

7.3. Developing a community science model

Schools can be understood as active drivers of cultural evolution, shaping mental models of the next generation towards selected societal aims. In the 21st century, these aims have increasingly shifted from the transmission of knowledge of facts, procedures, and basic literacy (though these are still important to achieve on a global level), to more integrated and progressive competencies such as interdisciplinary and critical thinking, systems thinking, cooperation, self-regulation, and ethical evaluation competencies, as illustrated by the range of

21st century competency frameworks that have been put forward (Griffin et al., 2012; OECD, 2019; UNESCO, 2017).

At the same time, schools and education systems are also challenged by the rapidly changing demands of the 21st century. These range from "curriculum overload" (OECD 2020) due to the ever increasing body of cultural knowledge that could be transmitted to the next generation in a finite amount of time, to the mental health issues of teachers and students such as burn-out, depression and anxiety that appear to stem from problems of how school culture and school climate are structured (e.g. Pelletier et al., 2002). Critically, these dynamics may be significantly interdependent, such as cases where curriculum overload may be implicated in teacher burn-out or retention.

Many frameworks and methods exist that aim to help schools in addressing these problems of practice - ranging from improvement of learning environments (e.g. Aldridge et al., 2012) and school climate (e.g. Thapa et al., 2013), enhancing teacher collaboration and collective efficacy (e.g. Donohoo et al., 2020; Kelchtermans, 2006), fostering student voice and participation (e.g. Beattie, 2012; Bell & Aldridge, 2014; Halliday et al., 2019; Mitra, 2004; Ozer et al., 2020), and fostering student social-emotional learning (e.g. Clarke et al., 2015; Durlak et al., 2011; Seligman et al., 2009).

However, this diversity of existing approaches to school improvement usually co-exist in a more or less isolated fashion, or as competing alternatives for improvement and evaluation in which to invest limited school resources. Education researchers point out that educators are increasingly overwhelmed by these competing solutions and frameworks (Kivel, 2015; Mintrop, 2016).

Furthermore, existing approaches usually do not explicitly integrate teaching and learning *about the science* underlying the theory and methods or the processes of behavioral and cultural change that such efforts target, especially not in terms of integration into school curricula.

With this context in mind, we have been working since 2019 within our Community Science Lab at the Department of Comparative Cultural Psychology to develop some foundations for a community science model within the context of interdisciplinary evolution and school improvement sciences. The scale of the challenges and opportunities in engaging this work, both in terms of theory and practice, are significant. For this reason, humility, caution, and room for exploration have served as guiding principles in our efforts to clarify this potential. Our approach has included both exploratory innovation development and theoretical synthesis work. We have co-designed youth-driven sessions with a core team of four local secondary school students, in 10th grade as of this writing, and we have been working with them since the 2019-2020 school year, their 8th grade year. In parallel with this, we have worked to synthesize a range of core theoretical and methodological perspectives from a diversity of disciplines across evolutionary anthropology and school improvement sciences.

The sub-sections below outline the core theoretical and conceptual elements emerging from our work in developing this model, prior to summarizing the case study of our Evolving Schools project.

7.3.1. Schools as field sites for community-based cultural evolution

Biologists and social scientists, including cultural evolution researchers, have long conceptualized the regions they are interested in as *field sites* for scientific understanding. Often this research has been of a top-down nature, with researchers driving the questions and reaping much of the benefits, relative to local communities (see Urassa et al. 2021). Trends in community-based research (e.g. Boyd 2014) suggests the need and possibility to advance more equitable and more effective scientific research through a range of participatory methods. David Sloan Wilson (2011) notably used evolutionary theories to conceptualize his academic home city of Binghamton, New York as a field site for cultural evolution, including school-based research collaborations (Wilson et al. 2011). Importantly, however, this approach lacked the conceptual learning integration we emphasize here, which we argue is central to an authentic commitment to community empowerment. In this context, we suggest that schools present a novel and important stakeholder context for advancing community-based methods in applied cultural evolution research, what we will refer to as *Community-Based Cultural Evolution (CBCE)*.

In contrast to a top-down view from external scientists onto the field site communities they study, CBCE suggests a more endogenous perspective. At every level of social organization (Fig. 1), individuals and groups, to varying degrees, can freely choose to think of themselves as an applied field site for understanding and improving valued outcomes. Additionally, members of such community-based field sites may variously choose to employ evolutionary concepts to advance their understanding of the cognitive, behavioral, and cultural change within their site. An approach or project can be said to be aligned with our concept of CBCE to the degree it is (1) endogenously and intentionally driven by stakeholders, (2) driven by the critical application of evolutionary concepts, and (3) engaged in iterative, cumulative, and participatory approaches to empirical understanding. These minimal criteria of course require further clarification and operationalization, however, here we will focus deeper on the implications of this model for school communities as related to aims in interdisciplinary evolution education.

Figure 1: The Multilevel Field Site*. From the individual to the planetary scale, members within these levels of organization can diversely conceptualize themselves as a field site for understanding and improving valued outcomes. In the context of schools, students and teachers can be empowered as community scientists to understand and influence positive cultural change within their school, local, or global communities.*

Schools are unique communities from this perspective, in that, school communities have the explicit aim, in one form or another, of cultivating cultural knowledge about the human condition and the nature of society while also cultivating a range of valued societal outcomes. Schools are communities in which the goals of conceptual understanding of cognitive, behavioral, and cultural change are deeply interdependent with the goals and experiences of change or stasis in these same domains of everyday life. Table 1, below, maps the differences that can emerge in a CBCE approach in which a school is focused on endogenous identity building around field site infrastructure and processes for improvement through conceptual understanding.

Table 1*: Differentiating field sites in applied cultural evolution versus Community-Based Field Sites in Community-Based Cultural Evolution.*

	Field site approaches to applied cultural evolution research	**Community-Based Field Sites / Community-Based Cultural Evolution**

Role of external Researchers	External researchers *identify the community* as a field site	Community members *drive identity-building* around field site *infrastructure* and *processes* for improvement
	External researchers conduct research *on* or *with* local communities	External researchers support design-based research *for* community-based field site development
	External researchers own the data and control workflows (possibly with community co-ownership)	*Community members own the data* and control workflows (possibly with external data sharing agreements)
Research foci	Primary focus is on *intervention development* and *testing*	Primary focus is on *field site development* (i.e. tools and systems infrastructure for multilevel intervention development and testing)
	The *conceptual understanding* of participants is not necessarily accounted for or focused on within an intervention	The *conceptual understanding* of community members is a central aim of field site development
	Weak focus on *structures of knowledge* in science and community (i.e. the focus is on the *expert development* of cultural evolutionary theories to empower locally adaptive solutions)	Strong focus on *structures of knowledge* in science and community (i.e. the focus is on *distributed social networks* of coherent and pluralistic cultural evolutionary theory development; see Hanisch & Eirdosh, this volume)

For that reason, understanding this interdependence between what students learn about humans and how they engage school or society as a human, is a keystone conceptual space for advancing CBCE as a model for school improvement, and therefore, a strategic focal point for self-identified school field sites to advance community-based research.

7.3.2. The Metacognitive Loop

We humans vary significantly in our *conceptual understanding* of human nature and human capacities for valued change or persistence (i.e. that is, our understanding of the *adaptive flexibility* of human behavior, cognition, and culture). We also vary significantly in those *actual capacities* for adaptive flexibility as individuals or groups. How variations in these two domains influence each other is less than clear. We refer to the likely complex and reciprocal interdependence of these domains as the *Metacognitive Loop* (Fig 2). Metacognition, a concept developed in the educational research of Flavell (1979), has been adopted and adapted across various fields of the human sciences, with subtle and more overt differences in meaning. Here we adopt a highly generalized concept of metacognition as *cognition of or about cognition*. This includes even perhaps unconscious self-awareness of our performance in a task (sensu Heyes et al. 2020), as well as more abstract reasoning about other cognitive agents or systems (sensu Boyer 2018). In this context, the concept of metacognition can be applied both to conceptual understanding and adaptive flexibility. That is, the conceptual understanding of an individual or group about the diversity and flexibility of human behavior, cognition, and culture, represents a form of *metacognitive knowledge* that may or may not be drawn upon in the (potentially adaptive) decision making processes of the agent(s) in a given situation. For example, conceptualizations of self or society as fixed, rigid, or inflexible, may reinforce individual or group behaviors that are maladaptive in some contexts. In contrast, conceptualizations of self and society as complex, dynamic and potentially flexible systems, may support the emergence of more adaptive behaviors (see Wilson, D.S. 2011; Ciarrochi et al. 2016). Interestingly, this example points to a potential overlap between a scientific (complex systems) understanding of human behavior, and adaptive everyday conceptual understandings that students may or may not develop.

Figure 2: The Metacognitive Loop. *A generalized and idealized conceptual model for highlighting the poorly understood reciprocal relationships between our conceptual understanding of human behavior, cognition, and culture, and our adaptive flexibility in these same domains. The model is intended particularly for the applied domain of general education curriculum design, to better engage questions regarding the role of human experience concepts in the curriculum.*

It could be argued that the holy grail of general education is to create the conditions to reliably support humans in developing a conceptual understanding of the human condition that is both scientifically adequate *and* adaptive towards identified values across different levels of social organization. Curricula are currently overloaded with unstructured and fragmented knowledge that may not reliably support these aims. We suggest that a deepened emphasis on multi-pedagogical approaches for the understanding of *human experience concepts* (Stern et al. 2021), the concepts of human *behavior*, *cognition*, *culture*, and *systems* that pervade the everyday lives of students globally, may frame one core opportunity for educational innovation (see also Hanisch & Eirdosh 2020c). That is, giving students ample opportunities to engage the human condition as conceptual *content* for learning, as well as *context* for creative and critical social change, may drive adaptive cycles of development within the metacognitive loop, and therefore provides a uniquely integrative framing for school improvement efforts. Such a direction requires the elaboration and operationalization of constructs adequate to the task.

7.3.3. Theories of Self, Schooling, and Society

As discussed in Hanisch & Eirdosh (this volume), cognitive scientists have advanced a view of human development as one of testing and elaborating theories about the world (Gopnik et al. 1999). Infants notoriously test and develop intuitive theories about the physical world around them (Shtulman 2017). Young children go on to develop theories of their social world, developing *Theory of Mind (ToM;* Wellman 1992) about the knowledge, beliefs, and goals of others. Parents of children around the world have been said to have adopted variant *ethnotheories of parenting* (Harkness & Super 1992) which reflects and governs their knowledge, beliefs, and goals in relation to their children. And all of us develop deeply intuitive theories or mindsets about ourselves and the societies in which we live, theories that influence and govern our most everyday and life changing decisions (sensu Wilson, T. 2011).

In this tradition, we propose that all humans who develop in environments where schooling is present will likely develop more or less intuitive *ethnotheories of schooling*. Some humans make their *Theories of Schooling (ToS)* explicit and testable, sometimes through more or less formal scientific methodologies (sensu Mintrop 2018), others may evolve their ToS through other individual or cultural learning processes. The myriad possible relationships between folk and expert ToS are not well understood, but likely significant for the design of adaptive learning environments. This chapter can not begin to fully explore the conceptual space suggested by ToS, but aims only to contextualize the concept within the broader CBCE model.

Specifically, we agree with Barrett (this volume) that, when thinking about folk ontologies, we should not apriori project or assume that such ethnotheories will fall within the same disciplinary structures as western academia (i.e. folk physics, biology, psychology, see Shtulman 2017). Instead, we offer a proposed structure of knowledge for understanding ToS (Fig. 3) that is specifically designed to support applications in curriculum design and school improvement.

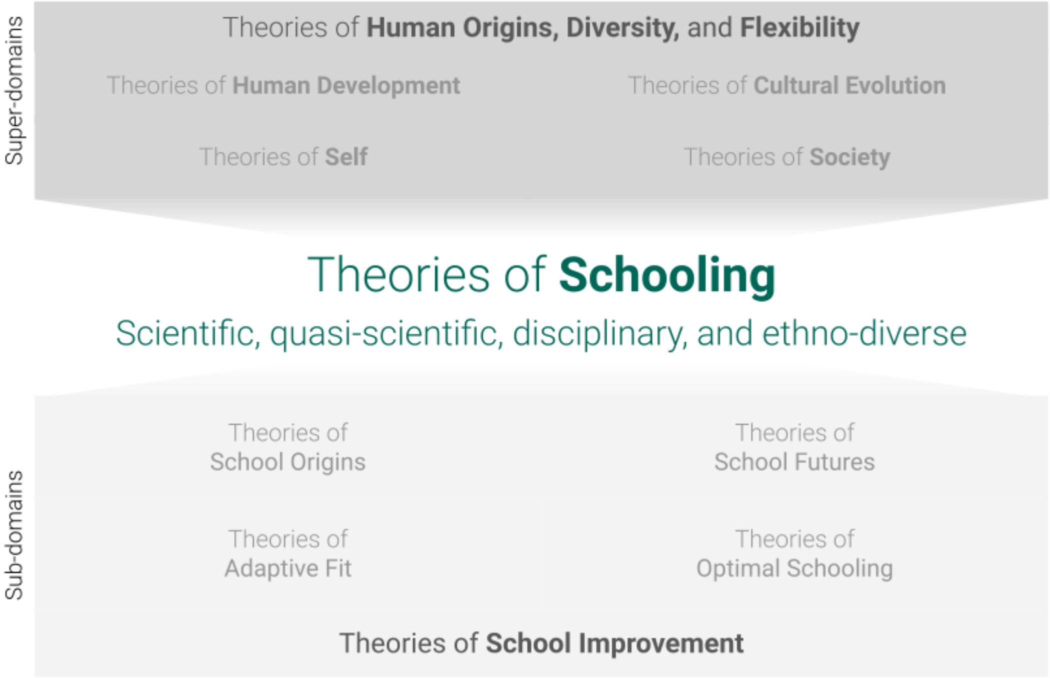


Figure 3*: A proposed structure of knowledge for the comparative study of scientific, quasi-scientific, disciplinary, and ethno-diverse Theories of Schooling.*

By understanding this structure of knowledge, curriculum designers interested in supporting informed participation of stakeholders in school improvement processes may begin to think about the deeper conceptual structures of the curriculum in relation to the development of scientifically informed and adaptive local theories of schooling. That is, what kinds of knowledge and transferable understandings do students have or need to adequately reason about the development of their own school? Stakeholder participation is about engagement of 'non-experts' or 'local-experts', and yet, the unique pedagogical aims of school communities might suggest we can look again at how best to scaffold student understandings (of themselves, schools, and society) that are adaptive for the world as we find or make it. That is, within a CBCE approach, we can seek to better understand, predict, and influence the (cognitive, behavioral, and cultural) evolution of adaptive theories of schooling across whole school communities (inline with Atkins et al. 2019).

A short-form model of this approach could be through an activity that allows students to make explicit their own theories of schooling in relation to a specific scientifically informed theory. We summarize exploratory lesson models from our Evolving Schools project in the next section, however, this exploratory work lacks integration within broader local school improvement processes, and thus falls short of our vision for a CBCE approach.

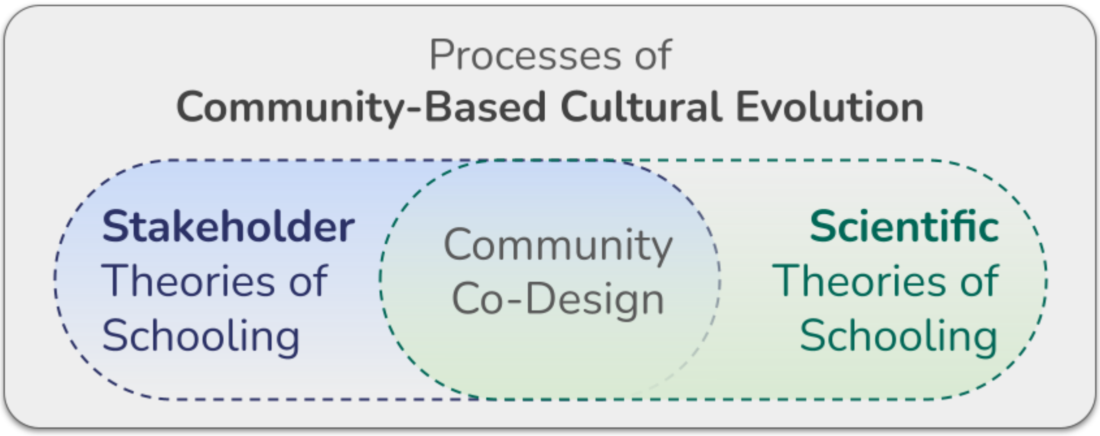


Figure 4: *Processes of Community-Based Cultural Evolution. The critical reflection of stakeholder ToS against diverse scientific ToS, can be one driver of community co-design of school improvement strategies. When these strategies then drive conceptual understanding and adaptive flexibility within and between individuals and school communities, processes of CBCE can be said to be occurring.*

A long-form model of CBCE would require a more systemic, interdisciplinary, and whole-curriculum approach. We suggest processes of *CBCE* can be said to be occurring within a school community when processes emerge that drive reflection between stakeholder and scientific theories of school improvement, in ways that can drive actual community co-design of on-going school improvement efforts (Fig.4).

Achieving this vision requires a more elaborated *theory of improvement* to support school improvement stakeholders in advancing this work locally.

7.3.4. Our Theory of School Improvement

First through our educational design work in the Global ESD project (www.GlobalESD.org), and now through OpenEvo (http://openevo.eva.mpg.de), we continue to refine our *Theory of Improvement* (Fig 5) derived from the concepts above. Critically, this is a minimal and generalized conceptual model of school improvement based on the aim of curriculum-scale coherence (see Hanisch & Eirdosh this volume).

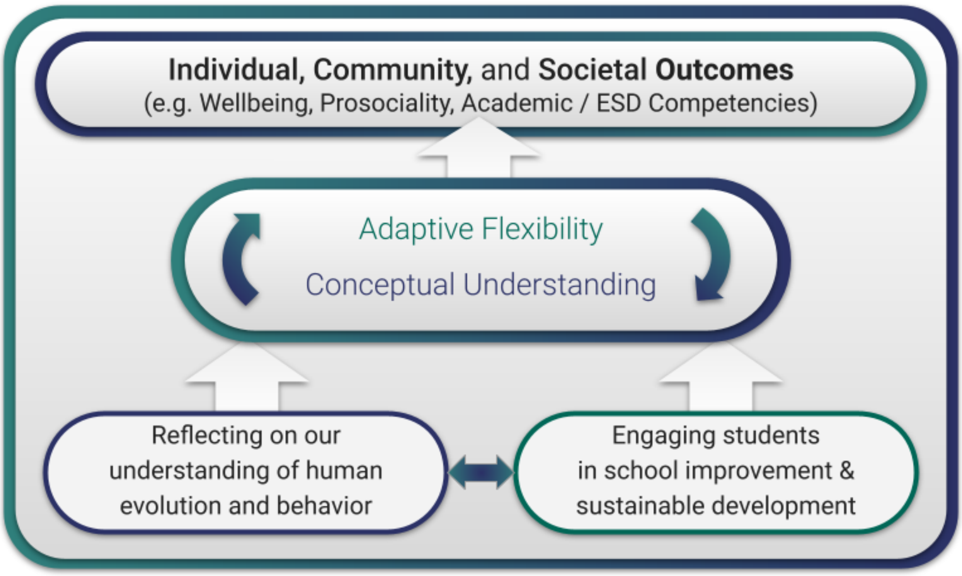


***Figure 5: The OpenEvo Theory of School Improvement**. A minimal and generalized model for organizing educational design research across grade levels, subject areas, and other aims or structures of the general education curriculum.*

The (very general) constructs and normative goals have been synthesized from evidence-based and widely accepted aims within global education discourse (see Hanisch & Eirdosh 2020c). The model emphasizes the complex, interdependent, non-linear, contextual, and reciprocal relationships between conceptual understanding and adaptive behavior (the *Metacognitive Loop*, above). Importantly, the model is not explicit about which specific conceptual understandings or which specific processes of adaptive flexibility are normatively optimal in a given context. Rather, the model focuses on engaging stakeholders in noticing if their current mental models, behaviors, or cultural institutions are adaptive in relation to their locally identified values, and maintaining or altering this variation as valued (sensu Atkins et al. 2019; Ciarrochi et al. 2016). This is a process which we suggest can be optimized through two broad, interdependent classes of curriculum and lesson scale targets of educational design, one more conceptual, one more contextual or experiential.

One class of intervention focuses on helping students reflect on their understanding of human evolution and behavior. This is the dimension of the model more focused on conceptual learning, yet also benefits from integrating diverse pedagogical approaches. To this end, we have advanced our *Educational Design Lab* model that develops tools and opportunities for students and educators to co-design interdisciplinary science and humanities teaching resources focused on understanding *human experience concepts*.

The other class of intervention focuses on engaging students in school improvement and sustainable development processes. This is the more experiential and contextual dimension, yet also requires deep conceptual learning. To this end, we have advanced our *Community Science Lab* model that develops tools and opportunities for students and educators to co-design community science projects that seek to understand and influence the cognitive, behavioral, and cultural variation of their school community.

Critically, both classes of intervention would theoretically be strengthened if a conceptually coherent framework for understanding human behavior, cognition, and culture can be structured (see Hanisch & Eirdosh, this volume). Such a structure is beyond the scope of this chapter, but we can provide a summary case study for exploring elements of our broader *Theory of Improvement* within our *Evolving Schools* project.

7.4. Case study: The Evolving Schools project

The core concepts for a CBCE approach outlined above developed through interdisciplinary knowledge synthesis and exploratory work within our student-centered Community Science Lab in Leipzig, Germany. Previous work (Eirdosh & Hanisch 2021) has documented the origins and early stages of the lab's *Evolving Schools* project, on which we will build to add context within the CBCE approach.

The Evolving Schools project started in March 2020 and has continued to develop through the present. The central guiding question for the project is:

How can students and school communities engage scientific perspectives on human behavior, cognition, and culture as a foundation for the participatory improvement of their own school?

Exploratory educational design work has led to a range of pilot projects engaging students in grades 7-12 in conceptual learning and critical reflection on evolutionary theories of teaching, learning, and schooling. The starting point for this work is often to elicit student conceptions, gaining a better understanding of how students' prior knowledge may inform their reasoning about a given topic. In this case, we want to develop methods for exploring the full potential landscape of ToS among students in general education contexts. Towards supporting research in this space, we have developed and begun to pilot a range of tools (Table 2) that enable the explication and documentation of some dimensions identified within the structure of knowledge outlined in Fig 3.

Table 2*: Patterns and pedagogical opportunities in everyday ethnotheories of schooling.*

Theories of Schooling Sub-Domains	***Evolving Schools* project toolkit examples**
Theories of **school origins**	Survey and classroom discussions on: Timelines for origins of schooling and teaching? Why was the first school created? Do other organisms engage in teaching like humans do?
Theories of **adaptive fit** between schools and students	Survey and classroom discussion tools on the *workability* of current school design elements

Theories of **optimal schooling & school improvement**	Student interview and focus group protocols on the adaptive value potential and scientific legitimacy of evolution-informed Self-Directed Education models of schooling

This early exploratory educational design work suggests that teaching and learning about the evolution of teaching and learning can be practically engaged within the general education curriculum, though much work remains to optimize and contextualize such activities locally. In this way, the integration and institutionalization of processes to help evolve adaptive ToS across stakeholders within formal school improvement processes should be a guiding aim in future developments.

This approach and the space of likely findings from such conceptual explorations do not lend themselves directly to short, simple, or single interventions, but perhaps point towards an opportunity for more interdisciplinary and integrative thinking about how students of all ages develop a conceptual understanding of human origins, diversity, and flexibility, as it relates to their everyday experience within their school community. The role of generalized evolutionary concepts in this understanding will depend on how students and curriculum designers more broadly understand the structures of knowledge in science and cognition. This in turn relates to how scientists within and outside of evolutionary theorizing understand such structures of knowledge. We suggest that the conceptually minimalist claims of generalizing evolutionary concepts (rather than “Darwinism” or “a Darwinian worldview”) provides a productive and coherent conceptual structure for helping students reflect on the nature of evolutionary concepts in educational contexts (see discussion in Hanisch & Eirdosh, this volume). That is, students can be better supported in understanding the complex causes (i.e. the variation producing processes and frequency changing processes) of their own evolving theories of schooling, in relation to the school they are a part of. Part of that support can include explicit explorations into the nature of evolutionary explanations across disciplines as part of broader Nature of Science learning goals.

Our Community Science Lab is now working to create a model open science workflow and digital research infrastructure to support secondary school students and school improvement stakeholders in working towards a truly global, open, community science project within the framework of Evolving Schools. Our hope is that this can be a complementary model for school communities to explore as they collectively evolve the content and context of their local curriculum.

7.5. Conclusions

In education, theory and practice are notoriously uncomfortable cousins in the family of school improvement practices. We need both of them at the table, but not everyone agrees about who should be running the reunion. Against that background, the conceptual complexities and clear lack of consensus on the generalizability of (or value of generalizing) evolutionary concepts, would seem to make interdisciplinary evolution a poor candidate for framing school improvement. Our suggestion, however, is relatively humble. Cognitive, behavioral, and cultural variation pervades the everyday lives of every human on earth. Schools are places for students to advance a conceptual understanding of that variation as well as the skills to adaptively influence the variation in their own lives and in their communities. Helping students to more explicitly reflect on their own conceptual models of this change in relation to the generalizability of evolutionary concepts provides a novel strategy for interdisciplinary science education and participatory school improvement research. In this context, any school in the world can self-identify as a field site for driving their own CBCE processes through local community science and interdisciplinary evolution education projects.

Our framing of CBCE as focused on the endogenous self-identification of a school as a field site is meant to frame a more explicit commitment of external research partners as having a limited but still direct role in creating the autonomy-supportive conditions for local communities to be (or become) effective agents of local (or global) change. Additionally, external research partners are central nodes in strengthening theoretical coherence and global knowledge exchanges across globally diverse school field sites. This means that there is a role for applied educational design researchers to support partner school communities with resources, processes, and infrastructure for more effectively driving CBCE processes. This also means there is a role of scientists from across all disciplines working to apply generalized evolutionary concepts, to better engage the international evolution education community towards supporting interdisciplinary evolution education resources that can drive the conceptual learning theorized to complement the broader school improvement aims of CBCE processes.

The conceptual landscape of CBCE represents a vast expanse of opportunities requiring further development. We invite all school stakeholders, students, teachers, parents, administrators, community members, and researchers across disciplines, to consider how you can support the emergence of a community-based field site within your local school or school system.

References

Aldridge, J. M., Fraser, B. J., Bell, L., & Dorman, J. (2012). Using a New Learning Environment Questionnaire for Reflection in Teacher Action Research. Journal of Science Teacher Education, 23(3), 259–290. https://doi.org/10.1007/s10972-012-9268-1

Atkins, P. W., Wilson, D. S., & Hayes, S. C. (2019). Prosocial: using evolutionary science to build productive, equitable, and collaborative groups. New Harbinger Publications.

Beattie, H. (2012). Amplifying student voice: The missing link in school transformation. Management in Education, 26(3), 158–160. https://doi.org/10.1177/0892020612445700

Bell, L. M., & Aldridge, J. M. (2014). Student Voice, Teacher Action Research and Classroom Improvement. (L. M. Bell & J. M. Aldridge, Eds.). Rotterdam, The Netherlands: Sense Publishers.

Böhm, G., & Pfister, H. R. (2015). How people explain their own and others' behavior: a theory of lay causal explanations. *Frontiers in psychology*, *6*, 139.

Boyd, M. R. (2014). Community-based research: Understanding the principles, practices, challenges, and rationale. In P. Leavy (Ed.), *The Oxford handbook of qualitative research* (pp. 498–517). Oxford University Press.

Boyer, P. (2018). *Minds make societies*. Yale University Press.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.

Ciarrochi, J., Atkins, P. W., Hayes, L. L., Sahdra, B. K., & Parker, P. (2016). Contextual positive psychology: Policy recommendations for implementing positive psychology into schools. *Frontiers in psychology*, *7*, 1561.

Clarke, A. M., Morreale, S., Field, C. A., Hussein, Y., & Barry, M. M. (2015). What works in enhancing social and emotional skills development during childhood and adolescence? WHO Collaborating Centre for Health Promotion Research. Retrieved from www.nuigalway.ie/hprc/%0Ahttps://aran.library.nuigalway.ie/handle/10379/4981%0Awww.nuigalway.ie/hprc/

Deci, E. L., & Ryan, R. M. (2011). Self-Determination Theory. *Handbook of Theories of Social Psychology: Collection: Volumes 1 & 2*, 416.

Donohoo, J., O'Leary, T., & Hattie, J. (2020). The design and validation of the enabling conditions for collective teacher efficacy scale (EC-CTES). Journal of Professional Capital and Community, 5(2), 147–166. https://doi.org/10.1108/JPCC-08-2019-0020

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. Child Development, 82(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Eirdosh, D., & Hanisch, S. (2020a). Can the science of Prosocial be a part of evolution education. Evolution: Education and Outreach, 13 (5). https://doi.org/10.1186/s12052-020-00119-7

Eirdosh. D., & Hanisch, S. (2020b). Education is an Evolutionary Science. Why Don't We Teach It That Way?. Essay In Eirdosh. D., & Hanisch, S. (2020). Evolution Education Without Borders: A Collection of Essays on Teaching Evolution as an Interdisciplinary

Science. Evolution: This View of Life. https://thisviewoflife.com/education-is-an-evolutionary-science-why-dont-we-teach-it-that-way/

Eirdosh, D., & Hanisch, S. (2021). Evolving Schools in a Post-pandemic Context. In *COVID-19: Paving the Way for a More Sustainable World* (pp. 465-480). Springer, Cham.

Engelbart, D. C. (1962). Augmenting human intellect: A conceptual framework. *Menlo Park, CA*.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906.

Frierson, P. (2018). Maria Montessori's metaphysics of life. *European Journal of Philosophy*, *26*(3), 991-1011.

Geary, D. C. (2005). Folk knowledge and academic learning. *Origins of the social mind: Evolutionary psychology and child development*, 493-519.

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (1999). *The scientist in the crib: Minds, brains, and how children learn*. William Morrow & Co.

Gray, P. The Evolutionary Biology of Education: How Our Hunter-Gatherer Educative Instincts Could Form the Basis for Education Today. *Evo Edu Outreach* 4, 28–40 (2011). https://doi.org/10.1007/s12052-010-0306-1

Griffin, P., McGaw, B., & Care, E. (2012). Assessment and Teaching of 21st Century Skills. In P. Griffin, B. McGaw, & E. Care (Eds.), Assessment and Teaching of 21st Century Skills. Springer. https://doi.org/10.1007/978-94-007-2324-5_2

Grobstein, P., Lesnick, A. Education Is Life Itself: Biological Evolution as a Model for Human Learning. *Evo Edu Outreach* 4, 688–700 (2011). https://doi.org/10.1007/s12052-011-0370-1

Halliday, A. J., Kern, M. L., Garrett, D. K., & Turnbull, D. A. (2019). The student voice in well-being: a case study of participatory action research in positive education. Educational Action Research, 27(2), 173–196. https://doi.org/10.1080/09650792.2018.1436079

Hanisch, S. & Eirdosh, D. (2020). Educational potential of teaching evolution as an interdisciplinary science. *Evolution: Education and Outreach,* 13 (25). https://doi.org/10.1186/s12052-020-00138-4

Hanisch, S. & Eirdosh, D. (2020). Causal mapping as a teaching tool for reflecting on causation in human evolution. Science & Education.

https://doi.org/10.1007/s11191-020-00157-z

Hanisch, S. & Eirdosh, D. (2020c). A Teacher's Guide to Evolution, Behavior, and

Hattie, J., & Yates, G. C. (2013). Visible learning and the science of how we learn. Routledge.

Sustainability Science. 2nd Edition. GlobalESD. http://guide.GlobalESD.org

Hanisch, S. & Eirdosh, D. (2021). Are humans a cooperative species? Challenges and opportunities for teaching the evolution of human prosociality. The American Biology Teacher, 83 (6). https://doi.org/10.1525/abt.2021.83.6.356

Harknes, S., & Super, C. M. (1992). Parental ethnotheories in action. *Parental belief systems: The psychological consequences for children*, *2*, 373-392.

Heyes, C., Bang, D., Shea, N., Frith, C. D., & Fleming, S. M. (2020). Knowing ourselves together: The cultural origins of metacognition. *Trends in cognitive sciences*, *24*(5), 349-362.

Kelchtermans, G. (2006). Teacher collaboration and collegiality as workplace conditions. A review. Zeitschrift Für Pädagogik, 52(2), 220–237. Retrieved from https://www.pedocs.de/volltexte/2011/4454/pdf/ZfPaed_2006_2_Kelchtermans_Teacher_collaboration_collegiality_D_A.pdf

Kivel, L. (2015). The Problem with Solutions. Carnegie Commons Blog. https://www.carnegiefoundation.org/blog/the-problem-with-solutions/

Kline, M. A. (2015). How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behavioral and Brain sciences*, *38*.

McKenney, S., & Reeves, T. C. (2018). Conducting educational design research. Routledge.

Mintrop, R. (2016). Design-Based School Improvement: A Practical Guide for Education Leaders. Harvard Education Press.

Mitra, D. L. (2004). The Significance of Students: Can Increasing “Student Voice” in Schools Lead to Gains in Youth Development? Teachers College Record, 106(4), 651–688. https://doi.org/10.1111/j.1467-9620.2004.00354.x

OECD. (2019). OECD Future of Education and Skills 2030. OECD Learning Compass 2030. A Series of Concept Notes. https://www.oecd.org/education/2030-project/contact/OECD_Learning_Compass_2030_Concept_Note_Series.pdf

OECD (2020), *Curriculum Overload: A Way Forward*, OECD Publishing, Paris, https://doi.org/10.1787/3081ceca-en.

Ozer, E. J., Abraczinskas, M., Duarte, C., Mathur, R., Ballard, P. J., Gibbs, L., … Afifi, R. (2020). Youth Participatory Approaches and Health Equity: Conceptualization and Integrative Review. American Journal of Community Psychology, 66(3–4), 267–278. https://doi.org/10.1002/ajcp.12451

Pelletier, L. G., Séguin-Lévesque, C., & Legault, L. (2002). Pressure from above and pressure from below as determinants of teachers’ motivation and teaching behaviors. Journal of Educational Psychology, 94(1), 186–196. https://doi.org/10.1037/0022-0663.94.1.186

Popp, J. A. (2012). *Evolution's first philosopher: John Dewey and the continuity of nature*. SUNY Press.

Ryan, R. M., & Hawley, P. H. (2016). Naturally Good? Basic Psychological Needs and the Proximal and Evolutionary Bases of Human Benevolence. In K. W. Brown & M. R. Leary (Eds.), The Oxford handbook of hypo-egoic phenomena. (Vol. 1, Issue May 2018). Oxford University Press.

https://doi.org/10.1093/oxfordhb/9780199328079.013.14

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. Hachette UK.

Seligman, M. E. P., Ernst, R. M., Gillham, J., Reivich, K., & Linkins, M. (2009). Positive education: Positive psychology and classroom interventions. Oxford Review of Education, 35(3), 293–311. https://doi.org/10.1080/03054980902934563

Stern, J., Ferraro, K., Duncan, K., & Aleo, T. (2021). Learning That Transfers: Designing Curriculum for a Changing World. Corwin Press.

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional science*, *32*(1), 9-31.

Sweller, J. (2008). Instructional implications of David C. Geary's evolutionary educational psychology. *Educational Psychologist*, *43*(4), 214-216.

Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology*, *4*(1), 147470490600400135.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A Review of School Climate Research. Review of Educational Research, 83(3), 357–385. https://doi.org/10.3102/0034654313483907

UNESCO. (2017). Education for Sustainable Development Goals Learning Objectives. UNESCO.

https://www.unesco.de/sites/default/files/2018-08/unesco_education_for_sustainable_development_goals.pdf

Urassa, M., Lawson, D. W., Wamoyi, J., Gurmu, E., Gibson, M. A., Madhivanan, P., & Placek, C. (2021). Cross-cultural research must prioritize equitable collaboration. *Nature Human Behaviour*, *5*(6), 668-671.

Van Bouwel, J., & Weber, E. (2008). A pragmatist defense of non-relativistic explanatory pluralism in history and social science. History and Theory, 47(2), 168–182. https://doi.org/10.1111/j.1468-2303.2008.00445.x

Wellman, H. M. (1992). *The child's theory of mind.* The MIT Press.

Wilson, T. (2011). Redirect: The surprising new science of psychological change. Penguin UK.

Wilson, D. S. (2011). The Neighborhood Project: Using evolution to improve my city, one block at a time. Little, Brown.

Wilson, D. S., Kauffman Jr, R. A., & Purdy, M. S. (2011). A program for at-risk high school students informed by evolutionary science. PLoS One, 6(11), e27826.

Wilson, D. S., Ostrom, E., & Cox, M. E. (2013). Generalizing the core design principles for the efficacy of groups. Journal of Economic Behavior & Organization, 90, S21-S32.

Ziadie, M. A., & Andrews, T. C. (2018). Moving evolution education forward: A systematic analysis of literature to identify gaps in collective knowledge for teaching. CBE Life Sciences Education, 17(1). https://doi.org/10.1187/cbe.17-08-0190

Chapter 8

Teaching for the interdisciplinary understanding of evolutionary concepts

Susan Hanisch[i], Dustin Eirdosh[ii]

Abstract

Evolutionary concepts are used, with varying and arguable degrees of scientific utility, across a wide range of disciplines. Evolution education, however, remains overwhelmingly within the confines of biology education, when it is taught at all within general education. The reasons for this disciplinary myopism are complex, and normative guidance for curriculum designers is scarce. This contribution explores how understanding the *structures of knowledge*, or the organization of facts and generalizations in science, cognition, and education, may help illuminate the educational potential and evidence-informed pedagogical practices appropriate for teaching about the interdisciplinary application of evolutionary concepts.

---

[i] University of Leipzig, Primary School Science Education; Max Planck Institute for evolutionary Anthropology, Department of Comparative Cultural Psychology; Friedrich- Schiller-University of Jena, Biology Education.

[ii] Max Planck Institute for evolutionary Anthropology, Department of Comparative Cultural Psychology; Friedrich- Schiller-University of Jena, Biology Education.

8.1. Introduction

Theoretical and methodological advances in evolution science suggest the possibility that evolution could and perhaps should be taught as an interdisciplinary science (Hanisch & Eirdosh, 2020a). However, such a potential generalization of evolutionary theory, as also explored in this volume, is perhaps among the scientific developments that most challenge the current structure of educational curricula and educational research and practice. While science is always advancing with emerging fields, theories, methods, and findings, and while curriculum development efforts often attempt to integrate these developments into school curricula, it appears that the generalization of evolutionary theory puts a particular strain on the structure of the educational system that has been forged over the second half of the 20th century.

Evolution education has been predominantly informed by the conceptualization of evolutionary theory known as the Modern Synthesis (MS). This framing of evolutionary change provided a core set of concepts and principles that have defined evolution education discourse and research, curriculum structure, materials, and assessment tools in the last decades, and presently. While it can be argued that these concepts and principles do provide a solid basis for some core understandings about how evolution operates in many cases, they may also present constraints in relation to broader educational goals.

In our educational design work, we regard a transferable understanding of evolutionary concepts and processes as a core set of learning goals for students to understand themselves, their fellow humans, their human-made world, as well as problems and solutions to sustainable development. In this work, we have been informed by three emerging fields of discourse: discussions around the possible value of an *Extended Evolutionary Synthesis* (EES), the emerging field of *Cultural Evolution Science* (CES), and the tradition of behavioral sciences known as *Contextual Behavioral Science* (CBS). To our knowledge, with some exceptions (e.g. Apodaca et al., 2019; Arújo, 2020; Pugh et al., 2014), the conceptualizations emerging from within and across these areas appear to be not currently part of the broader discourse on how to teach evolutionary science, particularly in secondary school and general education more broadly.

At the same time, evolution education continues to struggle with a range of persistent problems of evolution understanding and acceptance among students and the general public (Barnes et al., 2017; Gregory, 2009; Heddy & Sinatra, 2013; Legare et al., 2018; Pobiner, 2016; Rosengren et al., 2012; Sinatra et al. 2008). In Hanisch & Eirdosh (2020a), we argue that these persistent problems may be linked, albeit in complex and as yet not fully understood ways, to the persistence of *gene-centered* as opposed to *trait-centered*, interdisciplinary approaches to evolution education. That is, we suggest that defining the process of evolution solely in terms of changes in allele frequencies (as opposed to changes in trait frequencies) within a population presents significant constraints to solving the persistent problems of the evolution education field. Said another way, we argue that the constrained rather than generalized framing of core evolutionary concepts may be constraining our search for solutions to the challenges of evolution education.

Here, we build on this argument to clarify the role of structures of knowledge in science, cognition, and evolution education, and the relationships between them. This clarification suggests that critical reflection on the generalizability and contextually specific application of evolutionary concepts is a central yet underutilized pathway to deeper public understanding of evolution as an interdisciplinary science. Thus, teaching approaches that target the development of conceptual understanding and transfer of learning should take on a more central role in the evolution educator's toolkit. In Box 1 (p. XX), we take an excursion into the related field of complex systems science to see how domain-general concepts and processes of complex systems, and the learning goal of systems thinking, have already made their way into curricula. We argue that generalizable evolutionary concepts and the learning goal of evolutionary thinking can and should be equally considered as central in 21st century education.

In the sections that follow we unpack what is meant by *structures of knowledge* across the domains of science, cognition, and evolution education. We then highlight a range of implications for curriculum and instructional design of a generalized evolutionary theory.

8.2. Structures of knowledge in science, cognition, and evolution education

Overall, educators, scientists/philosophers of science, and cognitive scientists think of knowledge as structured, from concrete facts, events, examples or phenomena, to more and more abstract concepts, to hypotheses, generalizations, and principles linking several concepts, and finally a body of theory (fig. 1). In relation to this, educators, scientists and cognitive scientists also regard concepts, conceptions, and analogies as playing central roles in building structures of knowledge from existing prior knowledge. One of the indications of the interdisciplinarity of this view of knowledge, is the use of Bayesian causal inference models in both philosophy of science and cognitive science, which link causal hypotheses on different levels of generalization with inferences and evidence (e.g. Baraghith & Feldbacher-Escamilla, 2021; Gopnik et al., 2004; Gopnik & Wellman, 2012; Goodman et al., 2011).

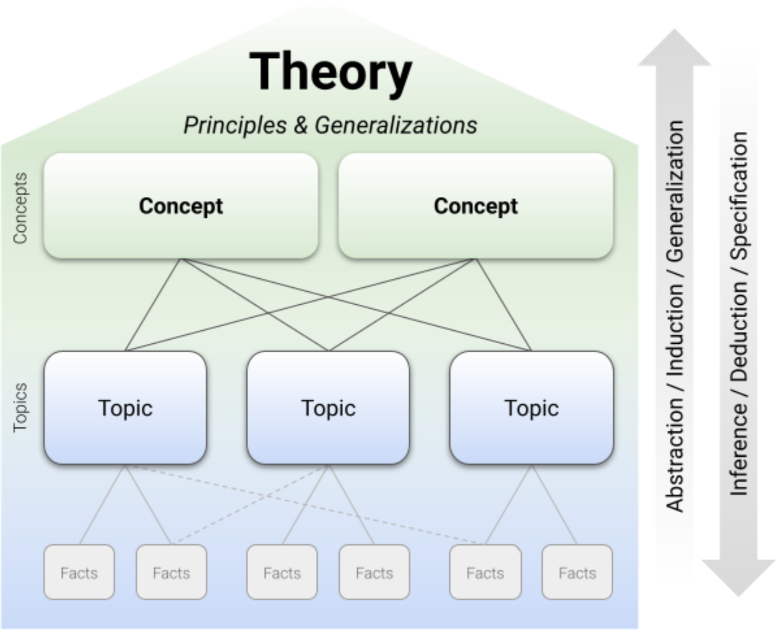


***Figure 1**: Basic structure of knowledge, highlighting the relationships between concrete facts, topics, generalizable concepts, generalizable principles and finally theory. Based on Erickson et al. (2017).*

8.2.1. Structures of knowledge in science

As the contributions of this volume show, the debate around a generalization or extension of evolution beyond the domain of biology often revolves around the assumptions and implications regarding the structure of evolutionary theory as a whole, and the role that different conceptualizations and applications of concepts and analogies can and should play in this. Because this issue is extensively addressed in the chapters in this volume (and elsewhere) we keep this discussion around SoK in science brief, and aim to only highlight points that are relevant for the following sections regarding SoK in cognition and education.

Scientific disciplines and philosophers of science vary in how they understand the structure of scientific theories. Overall, it appears that different disciplines or schools of thought differently value, strive for or consider achievable, *theoretical coherence* (across varying depths of explanation) and *scope* (across a breadth of phenomena) within their own field and in relation to other fields. Thagard (2007) argues that a theory that strives for “broadening coherence” (i.e. scope in terms of being able to explain more diverse phenomena) as well as “deepening coherence” (i.e. depth in terms of hypotheses being explained by more abstract theory) can be expected (but not guaranteed) to approximate (objective) truth the most based on what we know from the history of science and the structure of the world. Leaving aside such claims about “objective truth”, others have suggested pragmatic reasons for striving for coherence in terms of depth and scope. For example, regarding sustainability science, Tavoni et al. (2014) highlight how disciplinary silos stifle progress in addressing sustainability challenges if they lead to incoherent predictions about the effects of interventions, and authors call for unification and coherence across ecology, economics, and behavioral sciences. The field of contextual behavioral science also starts from the truth criterion of pragmatism, meaning a theory is “true” if it works in relation to a goal - in the case of contextual behavioral science, the pragmatic aim is to predict and influence human behavior to towards valued living with (theoretical) precision, scope and depth (Hayes et al., 2012). Further pragmatic reasons for coherence from the point of view of learning and education will be addressed in the sections below.

Related to the structure of knowledge is also the discussion around the role and value of analogies in science. In this regard, Stanley (2020) proposes that some of the disagreements in evolutionary science about the validity and value of an analogy between biological evolution and cultural evolution, are based on misunderstandings about the analogical transfer of evolutionary concepts and processes: some scientists seem to think that an appropriate generalization of evolutionary processes depends on the degree to which these processes are similar to processes of genetic evolution. On this view, any divergence in cultural evolution from biological evolution is seen as a weakness of claims about the "evolutionary" aspects of culture. Stanley (2020) argues instead that the appropriate analogy is actually between the abstracted concepts (see table below) and the diversity of domains to which the concepts are applied. That is, in the example table below (adapted from Eirdosh & Hanisch 2021), we should not be directly concerned with the similarities between genetic, cognitive-behavioral, and cultural evolution, but rather, we should focus on how pragmatically valuable it may or may not be to apply the abstracted evolutionary concepts in understanding each of the domains on their own.

***Table 1**: Analogy table highlighting some possible domain-specific instantiations of abstract evolutionary concepts in genetic evolution, learning, and cultural evolution.*

Abstracted Evolutionary Concepts	**Domains**		
	Genetic Evolution	Cognitive-Behavioral Evolution (Learning)	Cultural Evolution
How is **variation** of traits caused?	mutation, recombination	mistakes, recombination of prior learning, trial-and error learning, reactions to new environments, creativity, social learning	mistakes, recombination of ideas, trial-and error learning, reactions to new environments, creativity, between-group social learning
How does **selection** of traits occur	higher chances of survival and reproduction	selective attention, emotional strength, relation to prior learning, practical consequences	higher chances of survival and reproduction (*natural selection*); greater reward, appeal or attractiveness of the trait (*cultural selection*)

How are traits **inherited**, **transmitted**, or **retained**?	biological reproduction, mitosis/meiosis	encoding into long-term memory for later retrieval	social learning / imitation, teaching; technologies and infrastructure that endure

As Stanley describes:

> ... the ontologically minimalist process of evolution by natural selection can be realized by biological systems and by cultural systems, not because the two systems are alike, but because they both exhibit the relevant Darwinian properties of phenotypic variation, differential fitness, and heritability.
>
> The mistake here, or, at least, the misleading move, is the apparent attempt to model the cultural evolutionary mechanisms as being in close correspondence with the biological evolutionary mechanisms. These mechanisms don't have to be similar, or analogous, or even to correspond in some one-to-one like manner; the mechanisms can be substantially different.

As we build on Stanley's critical distinction in analogical reasoning, we will also here point to an important difference in wording that is particularly relevant in discussing the teaching of generalized evolutionary thinking. Stanley frames his analysis in terms of "Darwinian" principles and theory, and provides his (reasonable in our view) framing of what "Darwinian" means in this discussion. We suggest that in the applied domain of evolution education, we can make greater progress by focusing on the challenges and opportunities of teaching about *generalizable evolutionary concepts*. As this volume documents, there is not currently, nor is there likely soon to emerge, a significant singular consensus on what "Generalized Darwinism" is. In spite of this lack of scientific agreement at that level of theoretical organization, there is virtually no disagreement that many of the individual concepts employed in evolutionary explanations (e.g. variation, inheritance, selection, function, fitness, adaptation etc.) can be, and routinely are, employed in contexts beyond genetic evolution. The boundaries of when an explanation that invokes evolutionary concepts becomes an evolutionary explanation, or the degree to which an evolutionary explanation can be said to be "Darwinian", are fascinating and possible questions for classrooms to explore, but the existence of these questions are not arguments against the critical generalization of evolutionary *concepts* (fig. 2).

A focus on the level of concepts and an openness regarding a specific target body of theory also relates to the aspect of pluralism that some scientists and philosophers of science call for.

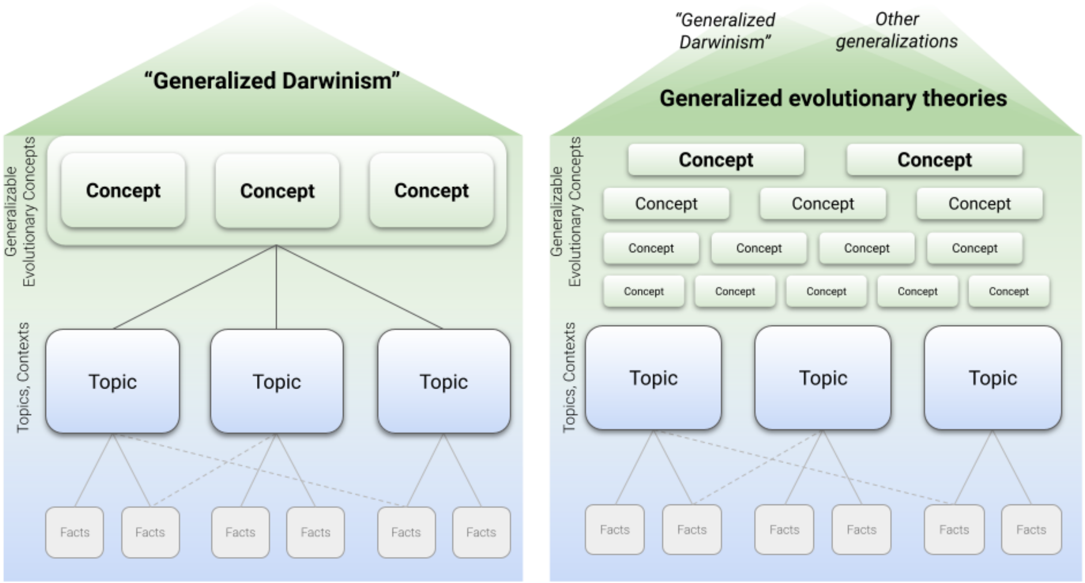


Figure 2*: A structure of knowledge culminating in a specific body of theory (e.g. "Generalized Darwinism") (left); and a more pluralistic structure of knowledge, allowing a complex landscape of generalized evolutionary theories to emerge from a range of evolutionary concepts and their relations.*

For example, Lohse (this volume) is skeptical of attempts to use cultural evolutionary theory to synthesize the social sciences, highlighting that there may be several legitimate reasons for the pluralistic nature of the social sciences. Van Bouwel & Weber (2008) propose an explanatory pluralism that is based on the plurality of questions that can be asked about any phenomenon, as well as the plurality of goals or purposes that an explanation is meant to serve. Their approach is non-relativistic because given a question and a purpose, several explanations can still be compared and ranked by their accuracy, adequacy and efficiency. Similarly, addressing the debate about an extended evolutionary synthesis, Baedke et al. (2020) highlighted how different levels of evolutionary explanation can fulfill different explanatory standards, including precision and idealization/abstraction.

As we will show in the following sections, this debate in science around coherence, pragmatism, and pluralism is contextualized by views about human cognition and learning as well as by the goal of education of helping students develop a networked (i.e. coherent), multi- and interdisciplinary (i.e. pluralistic), and helpful (i.e. pragmatic) understanding of the world.

8.2.2. Structures of knowledge in cognition

Psychology and cognitive science offer further insight into the role of structure of knowledge, concepts and analogies in learning and in relation to the potential of an interdisciplinary evolution education.

One example of this can be found in the field of psychology of science, which explores the psychological underpinnings of scientific reasoning, such as pattern recognition, categorization, association, causal reasoning, and analogical reasoning (Feist, 2006, 2013). Some developmental psychologists even make an analogy between the process of discovery and theory building in a scientific community and the process of learning during development (Gopnik et al., 1999; though the validity and usefulness of this analogy is also strongly debated, in some ways similarly to the debates surrounding the analogical nature of evolutionary concepts).

Regarding analogical reasoning, Gentner's structure mapping theory has been influential, which describes how learning and the build up of a structure of knowledge progress through analogical thinking:

> The process of analogical thinking can be usefully decomposed into several basic constituent processes. In a typical reasoning scenario, one or more relevant analogs stored in long-term memory must be accessed. A familiar analog must be mapped to the target analog to identify systematic correspondences between the two, thereby aligning the corresponding parts of each analog. The resulting mapping allows analogical inferences to be made about the target analog, thus creating new knowledge to fill gaps in understanding. These inferences need to be evaluated and

> possibly adapted to fit the unique requirements of the target. Finally, in the aftermath of analogical reasoning, learning can result in the generation of new categories and schemas, the addition of new instances to memory, and new understandings of old instances and schemas that allow them to be accessed better in the future. (Gentner et al. 2001 p. 9).

Generally, metaphor and analogy are considered to have central roles in human cognition and language. For example, Hofstadter (2001) proclaims that analogy is “the engine of cognition” and Lakoff & Johnson (1980) highlight how our everyday language is inherently metaphorical.

Of relevance for this discussion are suggestions regarding how humans tend to judge analogies to be “good”. For example, in the evaluation of analogies, studies show that factors like the degree of structural alignment; the amount of new knowledge that it generates; factual validity; adaptability of the relations to fit the target, and the relevance to current goals are influential (Gentner & Maravilla, 2018). Aspects of coherence as well as pragmatism and resulting pluralism are evident in these factors.

Regarding the degree of structural alignment, the systematicity principle has been proposed, which “reflects an implicit preference for coherence and predictive power in analogical processing” (Gentner & Colhoun, 2010, p. 37). Chesebrough et al. (2019) similarly highlight the role of coherence in efficient learning, stating that “Coherence exists when concepts “fit” together in ways that are unambiguous, consistent, and explicit. Content that is designed to optimize coherence creates vastly more effective learning.” and “Decoherence is created when the learner is unable to see clear connections, when the same concept is described in contradictory ways, or when the same language is used to describe different concepts”. This has clear implications for instructional design in evolution education, as will be highlighted below. Thus neural reward systems might indeed have evolved because of a higher adaptive value for coherence and generalizations in terms of scope and depth (Oh et al., 2020). This view has even led some cognitive scientists to speak of “explanation as orgasm” (Gopnik et al., 2001).

However, this doesn’t mean that human cognitive architecture always leads to optimized learning and approximation of objective “truth” (i.e. that humans are perfect Bayesian learners) or an entirely coherent structure of knowledge. This is partly due to various constraints, trade-offs, path-dependencies, and chance, just as in science and evolution. Furthermore, Legare & Shtulman (2018) highlight how humans seem to have coexisting domains of knowledge and pluralistic patterns of explanation which can be incoherent, such as both scientific and religious explanations, due to various sources of information and different kinds of goals as well as emotional aspects. Nonetheless, this does not mean that as educators, we can’t strive to help students develop a coherent structure of knowledge (see below) that helps them to integrate, e.g. everyday experience and scientific theories. Indeed, Legare & Shtulman (2018) propose integrated reasoning as one way to resolve previously incoherent coexisting explanations. Thus, similar to the debates in science, human cognition appears to be structured towards a capacity for both creating coherence and enabling pluralism.

Another important element for understanding the role of structures of knowledge in cognition is the role of prior knowledge and its relationship to new information. Haskell (2000, p. 10) stated that “All learning involves transfer from prior learning to a greater or lesser degree”, and Chesebrough et al. (2019) explain that “information that is consistent with an individual’s existing schemas is more quickly mapped onto neural networks in the brain where that information is stored, whereas information that is inconsistent with prior schemas requires more energy and resources for the brain to integrate“.

With regard to teaching evolution, one can thus ask to what degree different formulations of evolutionary concepts and theory can productively relate to existing student mental models about their biological, social and psychological world. Different answers to these questions have been proposed and explored in the evolution education literature (fig. 3-5):

- Student intuitive conceptions/theories as barriers to understanding;
- Student intuitive conceptions/theories as bridges to understanding;
- Student intuitive conceptions/theories as foundations for understanding.

We argue that answers to these questions depend on the presumed structure of evolutionary theory and the presumed definitions and scope of evolutionary concepts.

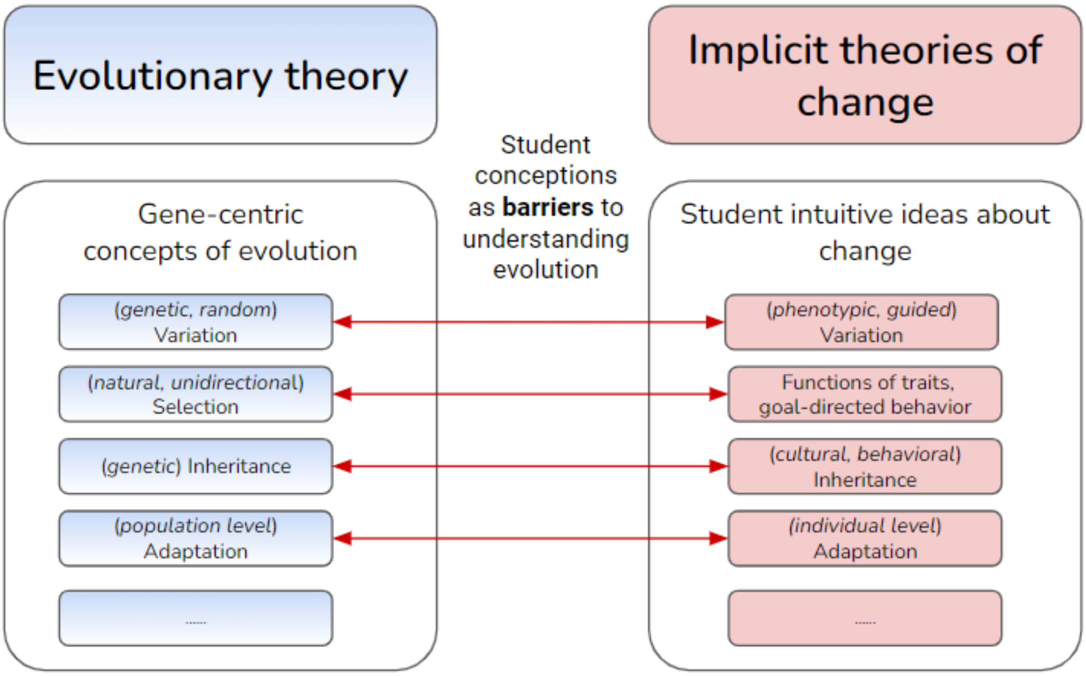


Figure 3*: Student intuitive conceptions/theories as barriers to understanding.*

On the one hand, one can presume that evolutionary theory and students' intuitive theories are inherently at odds with each other, that they cannot be integrated coherently and that therefore student preconceptions present barriers to understanding evolution, or at least that they need to be left aside in the evolution education classroom (Fig. 3). We argue that this position makes sense if one presumes strictly gene-centric conceptualizations of evolutionary concepts, as well as if one presumes that organism agency and behavior have no role in evolutionary explanations (both of which are common assumptions in evolution education, see below). After all, students do not experience genes in their everyday lives, and students experience agency almost constantly.

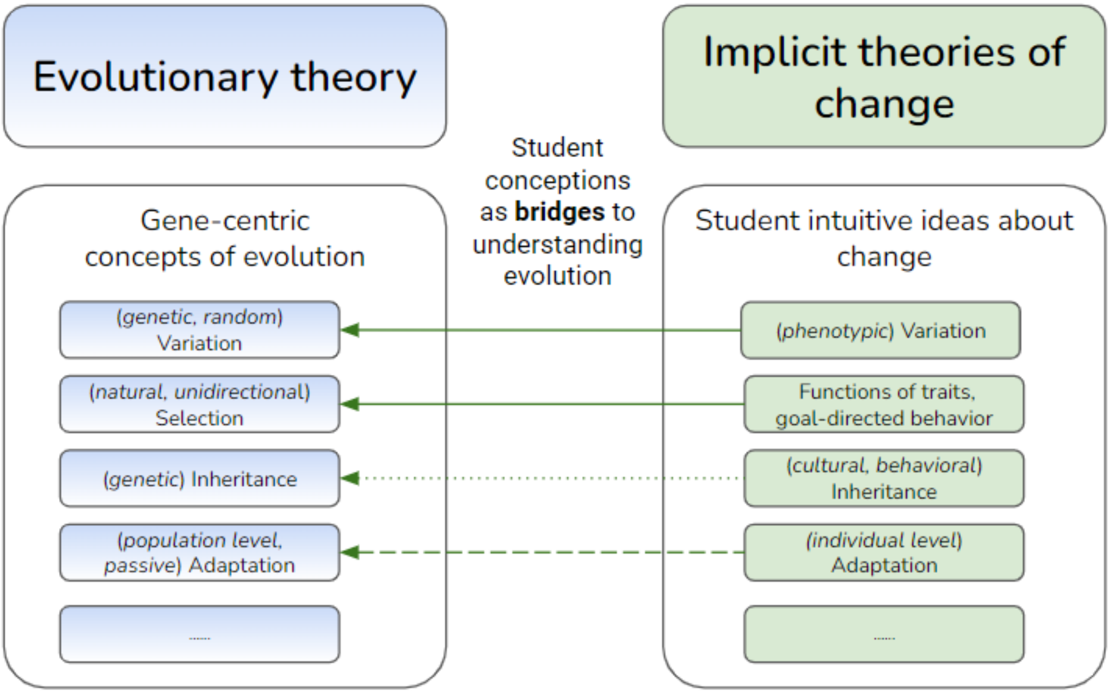


Figure 4: *Student intuitive conceptions/theories as bridges to understanding.*

Another view is that of students' intuitive theories of change as bridges to understanding evolutionary concepts (fig. 4). This is a view that has gained prevalence in education more generally, i.e. the idea that one cannot simply replace preconceptions with scientifically correct conceptions and that instead educators need to take preconceptions and prior knowledge as the only viable starting point for learning. In evolution education, students' intuitive understanding of function and needs has thus been regarded as a bridge or stepping stone towards understanding natural selection rather than a barrier (e.g. Bruckermann et al., 2020; Evans & Rosengren, 2018). Similarly, students' ability to see phenotypic variation is often used in evolution education in the early years as a starting point for understanding the role of variation in natural selection (see below). We are not aware of any studies in evolution education that explored students' intuitive understanding of (individual level) adaptation or of various inheritance streams (e.g. Moya et al., 2015) as bridges for understanding these evolutionary concepts, which is why in fig. 4, these aspects are presented as dotted lines. Importantly, most views of students' preconceptions as bridges towards understanding evolution still regard gene-centric and strictly defined conceptualizations of evolutionary concepts as the "target", whereby student conceptions are merely stepping stones and are to be "left behind".

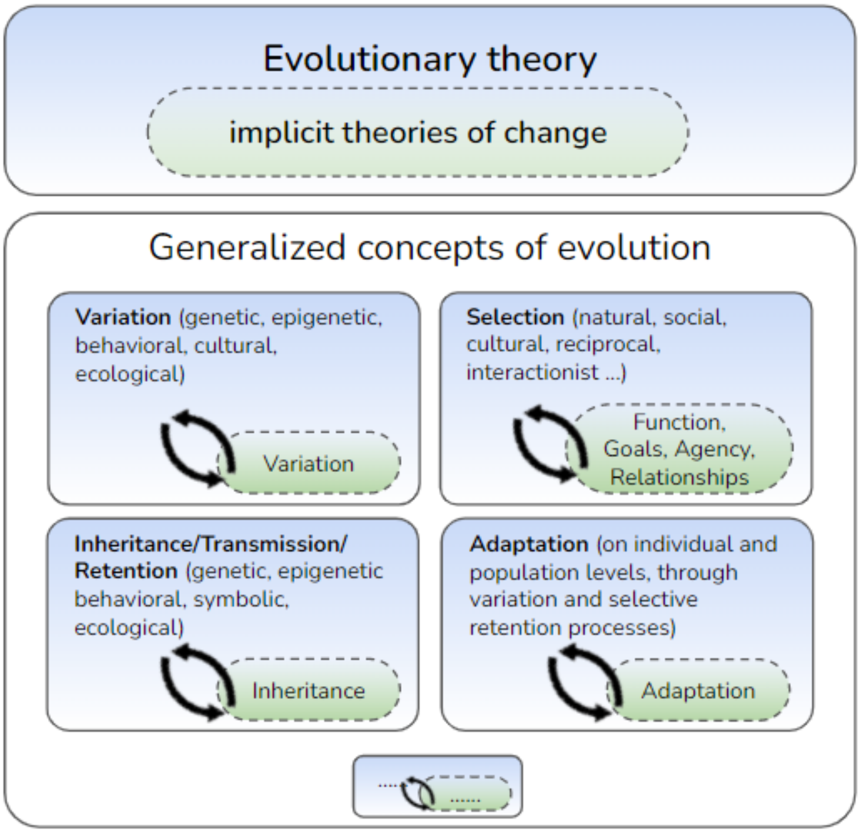


Figure 5*: Student intuitive conceptions/theories as foundations for understanding.*

A third possible view regarding the relation of intuitive theories and evolutionary theory is presented in fig. 5. Here we argue that evolutionary theory and concepts can be integrated with students' prior concepts, that student conceptions provide viable foundations for understanding. This is afforded by generalized conceptualizations of evolutionary concepts and processes which include, e.g. aspects like social learning as an inheritance mechanism, or which allow variations and changes in technology, in musical styles, or in individual development and learning to be explored as evolutionary phenomena. These can be productively *integrated* with students' existing mental models, while gene-centric conceptualizations of evolutionary processes cannot. Importantly, in the process, student intuitive understandings can also be re-represented and complexified, making them coherent with scientifically sound conceptualizations, such as regarding the role of decentralized causation in cultural change. For example, in the service of complexifying student understanding about goal-directedness, we can help them see that their own behaviors are often not intentional, or do not entail explicit goals, or create unintended outcomes, in order to build a schema of decentralized causation that is connected to their everyday experience and that they can use to understand decentralized causation and goal-directedness in evolution. This approach is in contrast to the currently predominant approach in evolution education which regards goal-directedness as barriers to understanding (represented in fig. 3), but, we argue, one that helps to build a conceptual coherence (scope and depth) and ultimately assists in deeper understanding.

Scientific reasoning is not just a purely rational process absent of emotional and motivational components. Cognitive scientists have also explored an affective and motivational dimension to analogical reasoning, in that certain analogies can elicit positive feelings, feelings of beauty and joy, enhance interest, inspiration or even self-confidence (e.g. Harrison, 2006; Thagard & Shelley, 2001). This aspect might have relevance to educational goals of fostering students' appreciation and motivation towards evolution, or enhancing attitudes like growth mindset and social-emotional learning. For example, in the field of contextual behavioral science, reinforcement learning is conceptualized as an evolutionary process (see table 1), and the self is conceptualized as a process, a context, a population or a system (Hayes et al., 2017). We would argue that such analogies present at least two sets of learning potentials to consider. One of them is that this schema of self as population, or self as complex system, instead of self as one fixed, essential entity, can be used to strengthen population thinking and a decentralized mindset (which should be of interest to the evolutionary biology educator). Additionally, the concept of the self as context or as process, instead of self as a fixed concept, relates to psychological flexibility (Kashdan & Rottenberg, 2010) and the learning goal of developing students' growth mindset (Dweck, 2012), i.e. the idea of the self as ever changing and able to improve through the ability for learning. In this way, by using generalizable evolutionary conceptualizations as the target of instructions, evolution education has the potential to build *both* understanding and emotional competency.

To conclude this section, besides the debate in science and philosophy, there are indications based on human cognition that tell us why we should indeed strive for coherence and encourage the flexible application of evolutionary concepts in developing students' structure of knowledge in (evolution) education.

8.2.3. Structures of knowledge in education

In this section, we aim to highlight how the structure of knowledge that evolution education presumes with regard to evolutionary theory informs educational standards, assessment tools, and materials. We suggest that this presumed structure also creates incoherence with respect to the framing of concepts and with respect to the application of evolutionary concepts to phenomena in students' lives.

Structure of knowledge and curriculum in evolution education

In education, perspectives around the structure of knowledge in science and cognition have strongly informed curriculum reform efforts since the end of the 20th century with the recognition that education needs to move away from coverage and rote learning of facts and topics and towards developing a deeper and transferable structure of knowledge in students (Erickson et al., 2017). Additionally, given the cumulative nature of scientific knowledge about the world, education systems are increasingly faced with a challenge of "curriculum overload" that demands a focus on core ideas that are transferable to a wide diversity of phenomena across domains (OECD, 2020).

Standards and curricula have also at least partly supported a more interdisciplinary coherence. For example, in the US, the Next Generation Science Standards have been developed for the STEM (Science, Technology, Engineering, Mathematics) fields (NGSS Lead States, 2013a). The framework identifies "seven crosscutting concepts that bridge disciplinary boundaries, uniting core ideas throughout the fields of science and engineering. The purpose of this framework is to help students deepen their understanding of the disciplinary core ideas, and develop a coherent and scientifically based view of the world" (NGSS Lead States, 2013b, p. 1). These crosscutting concepts include pattern, cause and effect/mechanism and explanation, systems, structure and function, stability and change. Similarly, in Germany, the biology standards delineate the core concepts of system, structure, function, and (individual, evolutionary) development (KMK, 2004). Thus, the influence of complex systems science (see Box 1) is prevalent in the structure of these natural science standards.

While these developments are laudable, we wish to highlight an important point that is also indicative of wider patterns within the traditional structure of school curricula. That is, we are not aware of an overall structure or curriculum standard that spans social and natural science domains in education (fig. 6). This has implications for generalized evolutionary theory. One interesting case in point is the rather ambiguous place of (human) behavior in the curriculum. In the US, the NGSS specifically excluded behavioral and social sciences from its definition of “science education” (National Research Council, 2012). On the other hand, the US Social Science Standards do not integrate biological explanations of human behavior (National Council for the Social Studies, 2013). Some German states have also developed curriculum standards that integrate the STEM fields on the one hand, and that integrate the social studies fields on the other hand (e.g. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur, 2014, 2016). Connections between (natural) science and social studies are encouraged, such as in the approach of socio-scientific issues (e.g. Zeidler et al., 2019), or within the NGSS in the core idea of science, technology and society, which prescribes that students explore “relationships among science, technology, and society”. (NGSS Lead States, 2013c). However, there does not appear to exist an overall structure of knowledge that integrates (human behavioral) concepts and theory across the natural sciences, social sciences and humanities, in current curriculum development. This ties into a much larger discussion about the role of social sciences or psychology within STEM education (e.g. Bray, 2010) as well as the philosophical possibilities and pitfalls of unification and synthesis across natural and social sciences (see also Lohse, this volume). The potential of generalized evolution to achieve at least partial theoretical coherence between natural and social science, and even humanities (e.g. through digital humanities) is thus currently not yet explored in curriculum development. Overall, 21st century human sciences are fundamentally interdisciplinary, routinely crossing natural and social science boundaries, and currently this appears to not yet be sufficiently reflected in curriculum reform efforts.

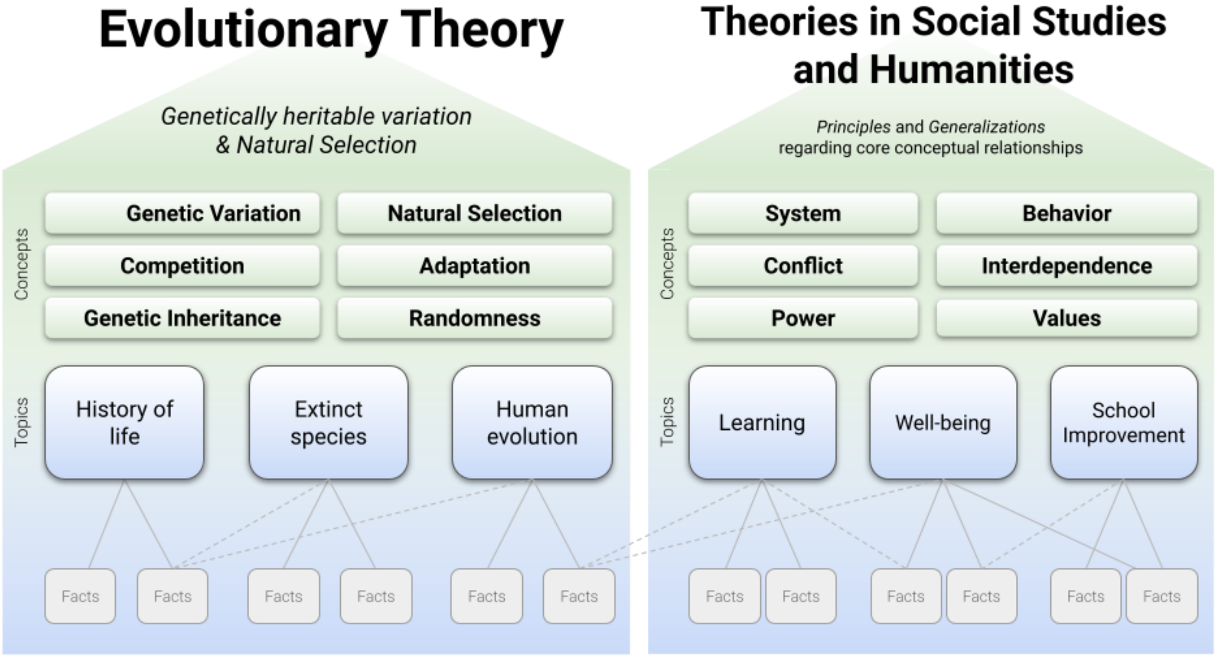


Figure 6*: Current structure of curriculum for many school systems around the world, with a distinct SoK for evolutionary concepts and theory which is largely, if not completely, separate from SoK in social studies and humanities.*

Specific learning progressions have also been proposed for evolution education. These implicitly or explicitly build on a notion of structure of knowledge, by identifying core ideas or principles as strands that are revisited over a range of age or grade bands in order to achieve increasingly complex, networked, and abstract levels of understanding. For example, the Understanding Evolution conceptual framework (University of California Museum of Paleontology, 2009a) is an influential framework in evolution education that was developed by a group of experts seeking to clarify and expand a learning progression of core evolution concepts for the K-16 grade levels. Other learning progressions have been developed specifically for the younger years (K-6; Lehrer & Schauble, 2012; Russel & McGuigan, 2019).

These learning progressions cut evolutionary theory at somewhat different joints (sensu Plato), or start with different *optimal structures* (sensu Bruner, 1974) of evolutionary theory. For example, the learning progression of Lehrer & Schauble (2012) for K-6 years is structured around the constructs of change (in individuals and populations), variation, and ecosystems, which they consider as ”serving as a conceptual foundation for reasoning about the theory of evolution later in their education” (p. 701); whereas other learning progressions do not consider understandings of *individual* change as relevant to the theory of evolution. If anything they might consider them instead as misconceptions that present barriers to understanding, which highlights the different approaches shown in Fig. 3-5 above.

What also seems to happen in learning progressions is that in the younger grades, evolution education standards and learning progressions start with very general understandings of concepts like variation, information, trait, and then progress towards an increasingly gene-centric understanding of these concepts (rather than, e.g. retaining the general conceptualization and exploring their different applications across domains). For example, in the NGSS life science standards on the core idea of heredity, the understanding about variation in the primary grades is “Different organisms vary in how they look and function because they have different inherited information.”, and in the following middle school grade levels, the general idea of information becomes constrained to genetic information, while in the high school grade levels, the role of environmental factors is again included, but it is not integrated or linked to the idea of inherited information (NGSS Lead States, 2013a).

We also observe other incoherent or changing use of concepts within standards and learning progressions. For example, the Understanding Evolution conceptual framework appears to reinforce a simple and direct genotype-phenotype relationship, or is at least not consistent about this relationship and its role in evolution. In the “Mechanisms of evolution” section of the framework, we find statements like “Evolution results from selection acting upon genetic variation within a population”; “Natural selection acts on phenotype as an expression of genotype”; “Phenotype is a product of both genotype and the organism’s interactions with the environment” (University of California Museum of Paleontology, 2009a). The framework also appears to be inconsistent regarding what levels of organization natural selection acts on. In the 12-16 grade level, we find the item “Natural selection is capable of acting at multiple hierarchical levels: on genes, on cells, on individuals, on populations, on species, and on larger clades.” while in another instance, it is stated that “Populations, not individuals, evolve.”(University of California Museum of Paleontology, 2009a). This is inconsistent or at least unclear, what is meant by “individual” and “population”. If natural selection is said to be able to act on multiple levels including genes and cells, how can it be that only populations (with the implication that what is meant is populations of individual organisms) but not individual organisms (as populations of cells and traits) evolve? While experts in evolution science may be capable of interpreting the intended logic behind these statements, this inconsistent framing of both the level on which selection acts and the very definition of the phenotype concept is very likely to drive confusion and provide obstacles to coherence and understanding, especially for novice learners. Helping students to develop a generalized and transferable understanding of the population concept (see Hanisch & Eirdosh, 2020b, for expanded discussion in evolution education; see Baragith, 2020, for a discussion in philosophy of biology) may help resolve this confusion.

Only at the 12-16 (post-secondary) grade levels is a definition of evolution offered in the Understanding Evolution framework, namely that “Evolution is often defined as a change in allele frequencies within a population.” (University of California Museum of Paleontology, 2009a). The term “often” is interesting in this regard, as it implies that there are also other - apparently non-problematic and scientifically valid, yet unidentified - definitions of evolution in use. We suggest that evolution educators should not just hint at these multiple definitions (indeed, multiple conceptions), but should explicitly engage students in reflecting on the similarities and differences between different conceptualizations of evolution, including everyday conceptualizations.

Structure of knowledge and instruction in evolution education

Looking at evolution education materials and assessment tools, we also find the clear influence of a gene-centric idealized formulation of evolutionary theory as well as instances of decoherence (*sensu* Cheseborough et al., 2019), unclarity and inconsistency regarding how to define and reason about evolutionary concepts. One might argue that the two are related - in that the gene-centric idealized formulation in fact leads to incoherence if it is applied to all phenomena uncritically.

A commonly used assessment tool is the Conceptual Inventory of Natural Selection (CINS; Anderson et al., 2002), which according to Mead et al. (2019) has been usd in 31 publications from 1990-2016. It targets a number of concepts through a number of context examples, such as Galapagos finches and the traits of beak size and shape; guppies and the trait of skin coloration; the evolution of three Canary Island lizard species from an African ancestor. There is a focus on morphological traits in the examples, but then multiple choice items include behavioral traits as if they can be reasoned about in the same way as morphological traits. On the concept of heritable variation, a distractor item is “Traits acquired during an organism’s lifetime will be inherited by offspring”; and on the concept of change in a population, a distractor item is “Learned behaviors are inherited”. The problem is that a significant constituent of biologists would agree that one type of trait, namely learned behaviors, are by definition “acquired” during an organism’s lifetime and are also often passed onto offspring in many species through the transmission mechanism of social learning (e.g. Hoppit & Laland, 2013; Jablonka & Lamb, 2005). This aspect is also something that students can be expected to have an intuitive understanding of based on their everyday experiences. Their reasoning (that learned behaviors can be inherited, or transmitted, to other organisms) might also be considered scientifically valid but such reasoning is not integrated productively into the educational practice but, in the case of this assessment item, rather thwarted.

Another prevalent assessment instrument is the Assessing COntextual Reasoning about Natural Selection (ACORNS, Nehm et al., 2012), which according to Mead et al. (2019) has been used in nine studies between 1990 and 2016. Among other things, it asks students “How would a biologist explain how [trait X] evolved?”, covering different context examples. We would argue that this question and the way that answers are analyzed reinforces notions that (a) all traits can be explained the same way, and (b) all biologists will explain a trait the same way. For example, student answers that say that they would need to know more about the function of a trait, are treated as misconceptions and receive a score of zero. It is quite astonishing that such a “blind”, even ritualized application of selectionist thinking across traits is put forward in such assessment and is considered an objective measure of “understanding”.

Overall, it appears that current assessment tools very often aim towards carving out and presenting an idealized structure of evolutionary theory (or sometimes targeting natural selection only) as if this is how biologists reason, generally and across contexts, about the evolution of traits, when in reality, the scientific discourse and practice is much more diverse and nuanced across the many biological and evolutionary subdisciplines. This relates to the debate around pluralism in science (see above). As Love (2013) points out, “reasoning in biological science is not homogeneous; biological science is composed of multiple perspectives that correspond to diverse explanatory aims and exhibit divergent reasoning styles. We must teach the heterogeneity of reasoning in biology”.

A gene-centric structure of evolutionary theory has also greatly influenced the research and development of evolution education materials and teaching strategies. In a review of research that may add to teachers’ pedagogical content knowledge on evolution education, Ziadie and Andrews (2018) found that topics such as the evolution of behavior, sexual selection and coevolution have received relatively little or no attention in educational research for undergraduate and secondary school biology teaching.

As an example, consider the chosen themes within the Teaching Evolution Through Human Examples project (Pobiner et al., 2018) - adaptation to high altitude, skin color, and resistance to malaria. All examples within this project cover morphological and physiological traits with a clear genetic basis, or which allow explanations of individual-level natural selection that do not entail more complex causation. Another unit within this project, which was not implemented in the study by Pobiner et al. (2018), was called “What does it mean to be human?”. In this unit, the focus was primarily on phylogeny, genetics and archeological concepts (Smithsonian Institution, 2015). In the final project of this unit called “Explaining Human Characteristics”, students are expected to provide evidence and create scientific explanations on the evolution of a chosen human trait from a list including increased brain size, a longer childhood, cooking, language, and the ability to create technologies (Smithsonian Institution, 2015, p. 47 ff.). However, important concepts like social selection, cooperation, social learning, niche construction, gene-culture coevolution, or important lines of evidence such as comparative, developmental and cross-cultural behavioral research, which are commonly invoked by evolutionary anthropologists to explain the evolution of such complex traits in our lineage, are not included in the unit purportedly asking “What does it mean to be human?”. Of course, it is possible to provide explanations only invoking mutations, morphology, and individual level competition and selection. To use the framework of van Bouwel & Weber (2008) - maybe such explanations can be considered the most *efficient,* but one can debate whether such explanations would be *adequate* for the purpose of asking the question “What does it mean to be human” in educational contexts.

Another example that illustrates how idealized concepts of evolution are implied to be adequate to explain the human condition, is a Massively Open Online Course (MOOC) developed by geneticist and evolution education researcher, Laurence Hurst, and colleagues (University of Bath, 2020). This MOOC is informed by Hurst's work in advancing the notion that to increase understanding (though not acceptance) of evolution, educators should "teach genetics first" (Mead et al., 2017). The online syllabus follows this logic, starting with the mechanistics and central tenets of a gene-centric idealized model of evolution. Then, in the final unit on "Human Evolution", learners are presented with archaeological evidence of the historical rise of humans, but not given any additional conceptual tools for understanding the multiple evolutionary streams of inheritance and cognitive-behavioral dynamics that are widely recognized as driving the evolution of our species.

A number of authors have emphasised that engagement and interest in evolutionary theory may be increased by pointing out to students the relevance of evolution to their lives. However, a gene-focused conceptualization of evolution constrains the examples that are often given regarding the relevance of evolution in everyday life and in society. For example, the Understanding Evolution conceptual framework states, regarding the relevance of evolution, "As with other scientific disciplines, evolutionary biology has applications that factor into everyday life, for example in agriculture, biodiversity and conservation biology, and medicine and health." (University of California Museum of Paleontology, 2009a). Similar themes are pointed out by many other authors (e.g. Pobiner et al., 2018). While these are important areas that can highlight to students how evolutionary biology is relevant for their everyday lives, important other areas such as understanding the evolutionary and developmental causes of human behavior, culture and cognition, are notably absent from this list. As we point out in Hanisch & Eirdosh (2020a), this may well have to do with the fact that many human traits do not have "simple" causes based on linear individual selection and genetics, meaning that they do not fit neatly into the idealized structure of evolutionary theory. They may even be considered as outside the realm of traditional STEM fields, and thus do not fit neatly into one of the traditional subject areas. Another application of evolutionary theory relevant to students' everyday lives is the role of evolutionary algorithms in technology and artificial intelligence, but again, given the gene-centric formulations of evolutionary processes, such a link is also largely absent in educational discourse.

Another educational approach that has been proposed to increase student engagement in evolutionary theory is the reflection on how humans might continue to evolve in the present and in the future. For example, Andrews et al. (2011) report on an intervention to teach natural selection through the question "Are humans still evolving"? The intervention sought to draw students' attention to the three necessary components of evolution by natural selection of trait variation, trait heritability, and differential reproduction. Within the intervention, these concepts were defined from a gene-focused conception, and these conceptions are then used to reason about the evolutionary change of traits whose distribution and spread may be caused through a much more complex set of factors and processes besides genes and natural selection through differential reproduction. Let us consider the example discussed in Andrews et al. (2011), on whether humans are *evolving* to become more obese, a trait that was proposed by students, presumably because they are aware of the increase in frequency of this trait in society. Looking at how student answers are interpreted (e.g. as misconceptions) and how classroom discussions are guided, the question on whether the change or spread of certain human traits can be considered the result of evolution (and therefore, whether students correctly apply evolutionary thinking to explain observed changes in trait frequency), depends crucially on how one defines concepts such as "inheritance" and differential reproduction or "fitness".

The following is an excerpt from a classroom discussion (Andrews et al., 2011, supplemental materials) to help students explore whether humans are indeed *evolving* to become more obese:

Instructor: "Is weight or tendency to put on extra weight heritable?"

Students: "Probably, but I don't know. I mean you see whole fat families, so probably it's genetic."; "Yeah, but families also all eat unhealthy or sit around all day, so maybe they just got fat because of that and not because of their genes."

(...)

Instructor: "Assuming it is heritable, do you think fat people are having more children than thinner people?"

Student: "Umm…no, I guess not."

The conclusion that students are meant to have drawn from this discussion is that humans are not evolving to become fatter, because on the one hand, it is questionable whether this trait is purely genetically inherited, and on the other hand, it is questionable that obese individuals have more offspring than other individuals. The discussion concludes with the question “What are some other explanations for why more people are obese?” (Andrews et al., 2011, supplemental materials).

Such classroom discussions may be more constructive for evolution education and enhance the development of a coherent structure of knowledge by, on the one hand, considering a variety of possible mechanisms of inheritance, thus reinforcing a transferable notion of trait transmission as important in changes of phenotype frequency, and on the other hand, by relating these concepts fruitfully to issues of public health in human populations, which would allow students to connect and reinforce their schema about evolutionary change. After all, possible trait transmission mechanisms other than genetic inheritance, such as social learning, were indeed pointed out by students in Andrews et al. (2011, see quote above) as possible explanations for the transmission or spread of obesity within a family. But these considerations were not considered to be relevant for an evolutionary account, and hence not viewed as relevant to the learning goals for this particular lesson activity, foregoing the opportunity to have classrooms engage in current public health issues while cultivating student understanding of evolutionary concepts in a more transferable fashion.

It is important to point out that we are not claiming that when exploring phenomena such as obesity, students should only be exposed to, or expected to generate, (generalized) *evolutionary* explanations, to the exclusion of other explanations, such as those that include other concepts from social sciences. As highlighted by van Bouwel & Weber (2008), “we should select the content of our explanation in such a way that it is adequate relative to our motivation for asking the question.” In this regard, van Bouwel & Weber (2008) also propose a “question-based pluralism”, meaning that “[f]or every social or historical phenomenon, there are many interesting and legitimate explanation-seeking questions that can be asked”, and hence different kinds of explanations can be considered adequate (or not adequate) depending on the question. In the case above, arguably the motivation for asking the question “Are humans evolving to become more obese?” was to see whether *evolutionary concepts* can play a role (and whether students can critically and adequately apply them) in explaining the observed phenomenon that the rate of obesity is increasing in society. Such focus on particular kinds of disciplinary concepts in a particular class is common educational practice and highlights that pluralism is somewhat baked into the curriculum. Clearly, in other subjects other concepts such as power, institutions, laws, norms, inequality, exercise, calorie, etc., may be equally applied to explore their role in explaining this particular phenomenon. When it comes to exploring *evolutionary concepts* in such phenomena, we argue that generalized (or generalizable) evolutionary concepts might make the discussion of such questions in the evolution education classroom more fruitful towards several learning goals, including in relation to providing conceptual coherence in evolutionary reasoning and increasing student motivation and interest.

Overall, despite the relevance of cultural evolutionary theory to students' everyday lives and its potential to connect to various educational goals, there appears currently an inconsistent approach to integrating cultural evolution into evolution education. For example, the Understanding Evolution framework introduced above does not contain any cultural evolutionary notions of concepts. At the same time, the website features teaching materials about cultural evolution and linguistics (based on Thanukos, 2008), where it is stated that "in fact, evolutionary concepts can be applied even beyond the biological world. Any system that has variation, differential reproduction, and some form of inheritance will evolve if given enough time." (University of California Museum of Paleontology, 2009b). A link is then made to the following conceptualizations in the framework: "Evolution results from selection acting upon *genetic* variation within a population." and "Evolution results from *genetic* drift acting upon *genetic* variation within a population" (emphases added). It is unclear how educators and learners should link the example content of the evolution of languages to the gene-centered conceptualizations put forward in the framework. Arguably the only ways to resolve this inconsistency is to either not include such examples as valid content for the evolution education classroom (thus forgoing opportunities for exploring evolutionary concepts across domains), or to amend the learning progressions to include more generalized conceptualizations of concepts.

At least in the German biology education context, the theme of cultural evolution appears in some state curricula (e.g. in the states of Saxony and Rhineland-Palatinate) and is covered in more or less depth in many biology textbooks. For example, in some German biology textbooks in the section on human evolution, we find paragraphs such as:

> Principally there are two mechanisms of transmission of information: genetic inheritance and learning from a model. (Jaenicke & Paul, 2004, p. 435, own translation)

> In contrast to biological evolution, cultural evolution enables the transmission of acquired traits. (Baack et al., 2016, p. 493, own translation)
>
> In many animal species, parents pass their acquired knowledge and skills on to their offspring - generating traditions and cultures. In humans the transmission of experience is especially elaborated. (...) The transmission of acquired traits from generation to generation, the imitation of behavior from models, learning and teaching - all this is summarized under the term of cultural evolution. Humans are influenced by it to a similar degree as by natural evolution. (...) Cultural and natural evolution have a number of similarities. Attractive, new ideas or fashions spread in populations with a similar dynamic as alleles. (Markl, 2018, p. 337, 338, own translation).

Students that learn about this generalized notion of evolution, might choose answers on standardized evolution understanding assessment tools (see above) that would be evaluated as “wrong” from a gene-centric perspective, such as responses concerning the heritability of learned behaviors.

Critically, the latter textbook quoted above (Markl, 2018) then goes on to reflect on the future of human evolution. In answering this question, the text falls back to a gene-focused definition of evolution: “Of course we are still subjected to the evolutionary mechanisms like mutation, selection, gene drift and gene flow through migration (...) Hence, the evolution of humans continues.” (Markl, 2018, p. 339, own translation). No further guidance is given in the book on how to navigate this changing use of the term evolution and the concept of trait transmission, or on how to reflect on the importance of cultural evolution in the future of our species. That is, no supports are given for students to construct a coherent structure of knowledge in relation to the interdisciplinary application of evolutionary concepts.

To conclude this section, we find in the field of evolution education a mismatch between the educational goals that evolution educators aspire to (which include transferable understanding *as well as* emotional and motivational elements and competencies) on the one hand, and the structure of knowledge that is being reinforced through the conceptual and instructional tools that are employed on the other hand. The presumed gene-centric structure of knowledge that influences curriculum development and instructional design in evolution education might in fact hinder the field from achieving the full scope of their goals.

Concurrently, we argue that the structure of knowledge (including its pluralistic nature) that is emerging from the current scientific discourse around generalizing evolutionary concepts and theory, may serve as a promising direction to address these issues.

8.3. Curriculum and instructional design implications of a generalized evolutionary theory

If we take seriously the goals of 21st century education regarding the development of deep and transferable understandings, and the role of structures of knowledge in science and cognition, we see a great potential of, even a need for, integrating interdisciplinary evolutionary sciences in evolution education.

In fact, an excursion to the field of complex systems science and its influence on curriculum design and instruction (Box 1), hints at the opportunities that the evolution education community could engage in.

In the following sections, we propose steps to enable the teaching of evolutionary concepts as generalizable concepts to be applied critically across disciplines beyond biology, similar to concepts of complex systems dynamics:

- Learning progressions need to be re-examined and re-designed with an emphasis on the development of transferable conceptual understandings of core evolutionary concepts across disciplinary contexts.
- Evolution education should more strongly embrace instructional methods of teaching for conceptual understanding and transfer of learning, and accordingly re-negotiate targeted learning outcomes and methods for their assessment.
- Evolutionary thinking can and should be seen and treated as a subset of (or overlapping with) systems thinking, and pedagogical practices aimed at teaching for a transferable conceptual understanding of systems science concepts should also be more strongly

integrated in evolution education towards understanding evolution science across disciplines.

Box: Excursion - Comparing complex systems science and evolution science in curriculum design

Complex systems science explores dynamics of systems across domains which are characterized by many interacting elements, and exhibit decentralized causality, nonlinearity, feedback loops, and emergence. This body of theory and methods was greatly advanced through developments in mathematics and computer science since the 1970’s and now informs disciplines like earth science, economics, and biology (Gleick, 1987). In a recent example of this influence, Jamie Davies has used a complex systems view of developmental biology to propose that, while the DNA double helix has become *the* icon of biology in the 20th century, a better icon for biology in the 21st century is the feedback loop: “The helix is too well-established an icon to be deposed any time soon. And yet, a simple loop would be a much more universal symbol of how life works at all of its scales and levels.” (Davies, 2014). This sentiment reflects a broader shift in the biological sciences towards the application of generalized systems concepts, as in systems biology (e.g. Noble, 2006), behavioral biology (e.g. Sapolsky, 2018), and recent approaches in genomics (Gregory et al., 2016).

Concurrently with these developments in scientific theory and practice, complex systems science has also informed the teaching of many different subject areas, and current curriculum standards increasingly recognize the need to develop systems thinking competencies in students. As a result, many education fields readily recognize that the science and understanding of systems dynamics can be applied to a wide range of phenomena. A whole area of educational science explores the educational practice and assessment for developing students’ understanding of complex systems, beginning in the primary school years (e.g. Booth Sweeney, 2006; Booth Sweeney & Sterman, 2007; Grotzer et al., 2017; Jacobson & Wilensky, 2006).

One might think that the generalization of evolutionary theory and of evolutionary concepts and processes to different domains would be similarly welcome by the education community. After all, evolutionary processes can be understood as a subset of complex systems processes, especially relevant to complex *adaptive* systems (see also Schurz in this volume). However, compared with systems science concepts, evolutionary concepts are not as commonly viewed in terms of interdisciplinary curriculum structures.

We propose that this difference in development may be for reasons that have to do with the different history and sociology of science that evolutionary theory has compared to complex systems science.

Despite the fact that evolutionary thinking has a long history across disciplines (see e.g. Veblen, 1898), the rise of the Modern Synthesis (MS) in the 1940's, which integrated insights of genetics and microbiology, advanced a popular view that evolutionary theory is properly understood as the purview of the biological domain, and by extension, the biology curriculum and classroom. In terms of complex systems science, however, the origin is rather in math, physics and computer science which dealt with complex systems dynamics in a much more abstract, domain-general fashion from the beginning (Gleick, 1987). Thus, developments in interdisciplinary evolutionary theory have different implications for biologists because, historically, teaching evolution has been the task of biology educators, who have learned a biology-centric conceptualisation of evolution, largely informed by the MS, and have developed curriculum standards, instructional methods and materials as well as assessment tools and research programs that target those conceptions.

Another complication of generalizing evolutionary theory in educational contexts, compared to complex systems science, is that evolution entails both microevolutionary processes of change, and macroevolutionary patterns of the history of life on earth, including age of life on earth and common descent of all species. Thus, in contrast to systems science, evolutionary theory also attempts to explain the origins of today's observable organisms and their traits, including human traits, with the help of the combination of microevolutionary and macroevolutionary components, and this has strong implications for our worldviews, values, politics, and understandings of our place in the world, and thus has a complex relation to normative claims. Thus, integrating generalized conceptual understanding of evolutionary processes into the curriculum may be conceptually more complex in some dimensions. However, this may not mean it is actually more complex to teach or learn in practice. This added complexity may in fact be viewed as a pedagogical opportunity rather than challenge.

Finally, evolutionary theory comes with much more political and moral baggage due to its history, compared to complex systems science, because evolutionary theory was advanced at a time when the socio-political landscape, together with the (comparatively low) level of knowledge about human diversity and its origins, lead to wrong inferences to the social domain, including eugenic notions. Eugenics being a field which many historical evolutionary scientists, including Darwin, Haeckel and Julian Huxley, can, from today's standpoint, be seen as engaging in an ethically questionable or unacceptable fashion (e.g. Fuentes, 2021, and responses to him). The notion of social darwinism implied that using evolutionary theory to explain human cognition and culture is tantamount to eugenics. Thus, educational systems in the second half of the 20th century effectively constrained evolutionary thinking to biology, and a whole generation of educators, in biology and other subjects, seems to have been trained to be wary of any application of evolution to the human domain, particularly human cognition and culture.

In summary, the potential and limits of generalizability of concepts in complex systems and evolutionary sciences seem to be similar if not identical, yet the role of these concepts in general education curricula has developed along two very different trajectories. This seems to have more to do with the different history and sociology of these two fields of science rather than their conceptual structure. We suggest curriculum designers and interdisciplinary scientific teams work together to rethink the potential of correcting these trends towards engaging students in the critical application of generalized evolutionary concepts.

8.3.1. Learning progressions and curriculum design

Current evolution learning progressions and standards are largely designed from a gene-centric (MS) structure of evolutionary theory. While this may have been functional in the past because evolutionary theory has been dominantly situated in the biological domain, we find that the various generalizations of evolutionary theory within current scientific work (as well as the inherently metaphorical nature of evolutionary concepts such that we find them in everyday language) demand new frameworks that explicitly incorporate more generalized notions of evolutionary concepts and processes within biology (see e.g. Araújo, 2020) as well as across the general education curriculum (Fig. 7).

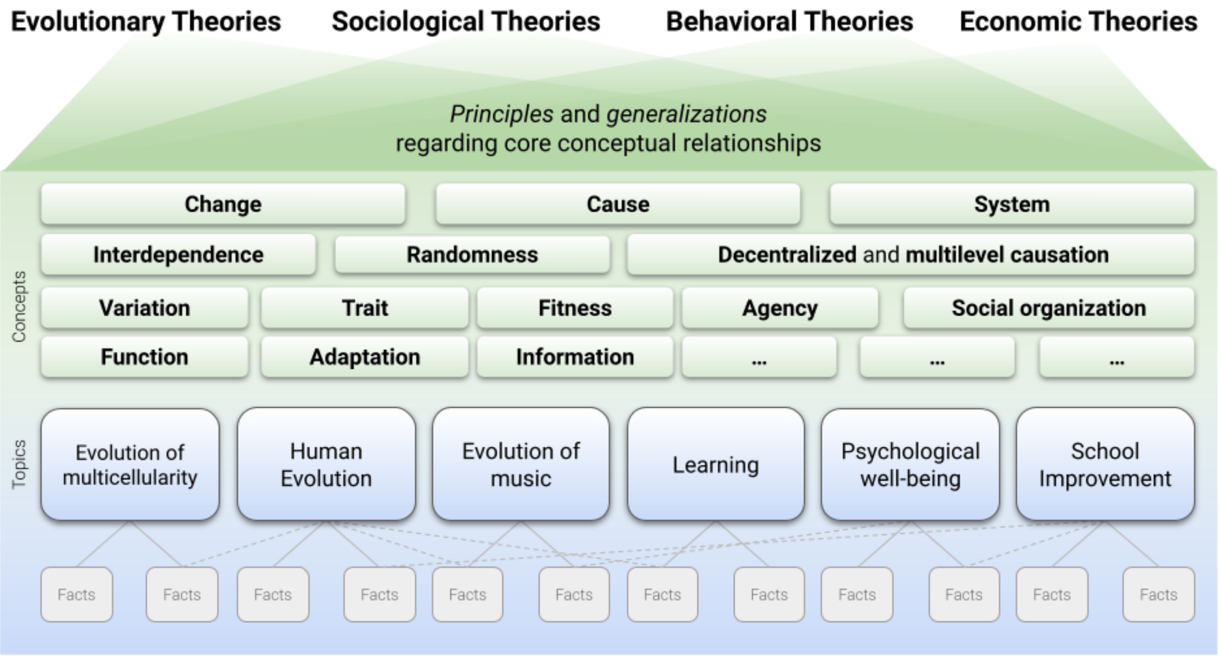


***Figure** 7: Possible structure of curriculum with the role of evolutionary concepts (besides other disciplinary concepts) across natural and social science theories.*

Above we highlighted how, in current learning progressions, understandings in early years tend to start with generalizable conceptions (e.g Lehrer & Schauble, 2012; University of California Museum of Paleontology, 2009a), while in later years these conceptualizations become increasingly gene-centric. We propose that learning progressions can instead focus on the generalizability of concepts and principles and their context- or domain-specific conceptualizations throughout. Evolution education can renegotiate its scope in the curriculum, and go beyond gene-centric notions, allowing connections to be made to a wider range of phenomena that include culture, psychology, and technology (fig. 7).

However, this requires a more coordinated effort across traditional domains in education systems, especially across the natural and social sciences. Biology educators might object that they do not need to concern themselves with a generalized evolutionary theory, such as cultural evolution, in their teaching of evolutionary concepts. We argue that this stance invites incoherence in the curriculum and in student cognition and results in inefficient learning of evolutionary concepts. Scientific developments like cultural evolution have already entered into the curriculum (see examples above), but under the currently prevalent framings of evolutionary theory in standards, assessment tools and materials, this increasingly creates incoherence and confusion. A coordinated approach to curriculum development that spans the natural and social sciences would offer a great new frontier for coherence, expanding on the curriculum reform movements that aim to strengthen interdisciplinary connections.

8.3.2. Teaching for conceptual understanding and transfer of learning

Informed by cognitive science and the structure of knowledge, many educators highlight the role of teaching for conceptual understanding, analogical reasoning, and transfer of learning in order to develop deeper structures of knowledge and enable students to use their understandings in novel contexts.

We argue that the diverse generalizations and applications of evolutionary concepts in science as well as the philosophy of science discussions around these developments, provide great opportunity for a renewed emphasis on teaching for conceptual understanding and transfer of learning in evolution education. In this regard, Reydon (2020) criticized the conceptual change literature for assuming that there are correct and consensus understandings of scientific concepts that are the target of instruction, including in evolutionary science. He highlights how scientists often or usually use different conceptualizations of a concept, and that conceptualizations change over time. As the debate around a generalization of evolutionary theory shows, there is no singularly true “consensus” view on what concepts of evolutionary theory mean and to which phenomena they can be applied, and thus any particular understanding of a concept (Reydon uses the example of fitness), should not be the target of instruction. Instead, he argues that educational practice should help students in engaging in a whole space of how a particular concept can be understood and conceptualized by exploring the history of science, their own everyday conceptions, or conceptions of scientists studying different phenomena. In this regard, it is interesting that evolution educators and learning progressions often highlight and include the role of teaching the Nature of Science as part of fostering evolution understanding (e.g. Nelson et al., 2019; University of California Museum of Paleontology, 2009a). However, while Nature of Science is usually thought to include this view implicitly, it does not seem to include teaching about the *nature of concepts* and *conceptions* as part of nature of science and scientific discourse explicitly (see e.g. Lederman et al., 2002).

Erickson et al. (2017) and Stern et al. (2017, 2021) propose a range of approaches to focus curriculum and instructional design on developing such understandings of concepts and generalizations. The learning transfer method proposed by Stern et al. (2021; fig. 8) targets conceptual understanding, including the metacognitive understanding of concepts themselves. It starts by helping students attain understandings of certain core concepts of a field by exploring what they know about the concept, finding examples and nonexamples, identifying attributes and constructing their own definitions. Over time and by exploring further examples of a concept, students also deepen and complexify their understanding. As a next step, teachers help students explore how concepts relate to each other to form generalizations and principles by investigating specific phenomena where these relationships can be "uncovered". Finally, teachers help students transfer their understandings by exploring how those generalizations and principles apply (or don't apply) across various phenomena, in the process further complexifying their thinking.

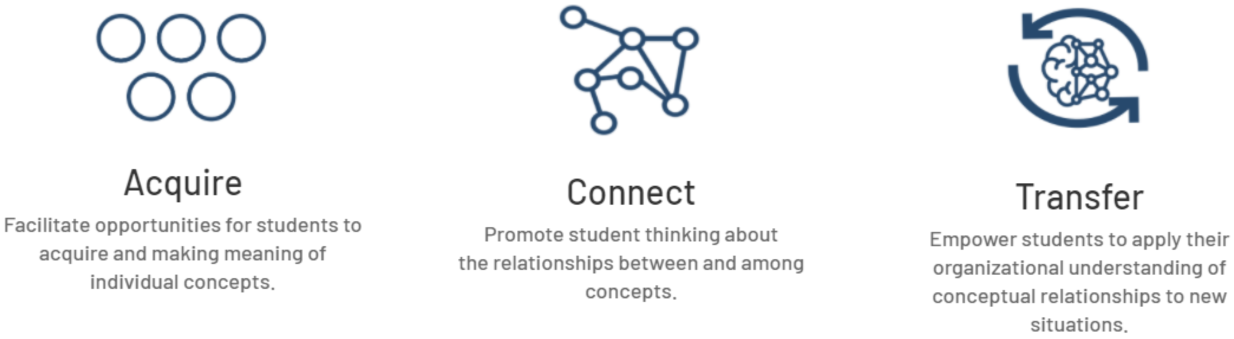


Figure 8*: The basic process of teaching for transfer (Image source: based on Stern et al., 2021, p. 10).*

A large body of literature also focuses on instructional methods for the use of models, case comparisons, and analogies for targeting transfer of learning (e.g. Alfieri et al., 2013; Haskell, 2000; Harrison & Treagust, 2000; Vendetti et al., 2015). The instructional method of analogy mapping guides students in comparing phenomena by underlying principles or concepts, as well as differences between them, enabling them to look beyond surface features and achieve a more abstract representation (see table 1; Glynn, 2008). Interestingly, far analogies (that are more different in surface features) have been shown to lead to deeper learning than near analogies (Walker et al., 2018).

With such tools, we can help students re-represent and complexify their existing mental models about change, in order to make them coherent with evolutionary and decentralized, complex change (see Fig. 5). We would first help students unpack, complexify, explain, and relate the existing conceptualizations they might have in their mind about particular concepts such as "adaptation" or "development" or "learning", to construct a definition of these terms, to find examples and non-examples of these concepts, or to create a causal diagram linking environment, organism behaviors, other traits, genes etc. (see Hanisch & Eirdosh 2020c). As students then explore adaptation on phylogenetic time scales, we do not throw out their existing mental models about the concepts, but we would encourage them to compare the two - how are they similar, how are they different, how does understanding one of them help us in understanding the other, or how does understanding one change and enhance our understanding of the other, how does comparing both lead to a more abstract, transferable schema about the nature of change by evolutionary processes?

Students can even be offered conceptual questions such as “How is evolution like individual development, how is it different”? or “How is evolution like learning, how is it different”?, or “How is the evolution of species like the evolution of [cultural trait X], how is it different?” as an anchor question that is revisited throughout a unit on evolution (see e.g. Pugh et al., 2014). Such explorations can be adapted for different ages or classroom contexts.

A focus on conceptual understanding also calls for new approaches to assessment. Stern et al. (2017, 2021) propose assessment tools including self-and peer-assessment that encourage students to reflect on their changing conceptualizations and apply them critically to new and increasingly different phenomena. One can conceive of final assessment tasks in which students are presented with a phenomenon of trait change in the world, such as the spread of a new virus variant, or of anti-vaccine sentiments, the adoption of a technology, or changes of the distribution or characteristics of species and ecosystems in response to climate change, and to use their understanding of evolutionary concepts and processes to explain these phenomena and even develop or reflect on potential interventions. This would be quite similar to the question “How would a biologist explain…” as used in one evolution assessment tool (Nehm et al., 2012; see above), but reframed more generally as “How would an evolutionary scientist explain …”, and it would look for students to ask the right questions and look for the right evidence in relation to the phenomenon, rather than for a rigid application of words and concepts. On the other hand, educators often still need efficient and standardized assessment tools such as multiple choice tests. Such assessment tools that are currently in use need to be reexamined critically to assure that they do not contain “trick questions” that leave space for interpretation (especially regarding the different possible conceptualizations of evolutionary concepts), are not merely testing for the ritualized learning of “how to pass the test” or “what the teacher wants to hear” and do not negatively assess students whose thinking can in fact be considered scientifically valid.

8.3.3. Integrate and foster systems thinking in evolution education

As we have argued in Box 1, the teaching of evolution as a generalized and interdisciplinary science can be informed by the success of teaching about domain-general complex systems dynamics, even at primary and middle school levels. Various teaching tools such as causal maps (Hanisch & Eirdosh, 2020c) and computer simulations (e.g. Centola et al., 2000) are available to teach students the complex nature of evolutionary change, from the start, with relatively simple models. For example, Roberts (1978) showed that middle school students could be taught to understand complex systems dynamics on a level comparable to MIT undergraduate courses with the help of scaffolded causal diagrams and explanation prompts. Thus, from a pedagogical view, it is important to emphasise that leading with complexity in the evolution classroom does not need to be "too complex" for students at certain grade levels given the right teaching methods.

We argue that evolution education can and should more strongly integrate such instruction and assessment methods of complex systems education into the evolution educators' toolkit. For example, causal diagrams as used in complex systems education can be modified to include mechanisms and causal factors that are considered relevant in the evolution (and development) of particular traits, populations or systems of interest. In Hanisch & Eirdosh (2020c) we propose such a causal mapping toolkit for evolution education that allows the integration of development and evolution, of a range of causal factors (including behavior, culture, social environment as well as genes) and processes, complex systems dynamics such as feedback loops and decentralized causality.

We can also use instructional tools that reduce complexity by providing scaffolded representations and examples that chunk complexity *at the right joints*, or sequence it appropriately. In this way, specific causal relationships between factors and causal mechanisms can be introduced sequentially, e.g. by exploring relevant phenomena that help students uncover those relationships. The individual-level, unidirectional natural selection of morphological traits that have relatively direct genotype-phenotype relationship is *just one* of those phenomena that can be explored (Fig. 9a). However, importantly, we argue that evolution education should not stop there. More complex traits such as behavioral traits that include the important role of learning or of the social environment, can equally be explored and visualized, so long as they are integrated into an overall coherent structure of knowledge, rather than treated as "a new topic" (Fig. 9b). Diverse sources of facts, including genetics and archeology, but also examples from developmental, cross-species, cross-cultural behavioral and social science research can help to illustrate and uncover those relationships. Students can and should be challenged to articulate and defend their causal models on the basis of accessible information.

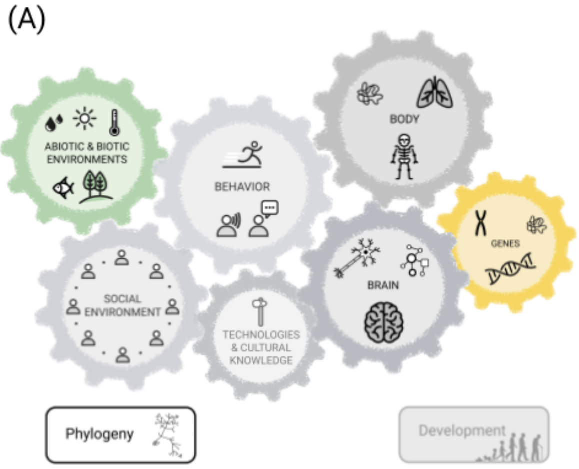
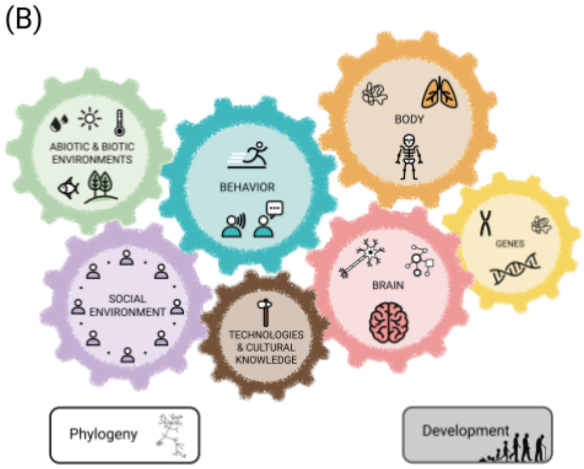



***Figure 9**: A possible "cutting at the joints" of complex causal relationships in the evolution and development of developmental systems. (A) Emphasis on unidirectional natural selection of genetic variants by a particular environment over phylogeny, while backgrounding backgrounding other factors and causal relationships, which is helpful when exploring particular evolutionary phenomena; (B) Expansion of the model by adding further traits and factors, thus building a more complete mental model of the causal relationships that need to be considered in explanations of the evolution and development of traits and developmental systems.*

8.4. Conclusions: Evolving an interdisciplinary evolution education?

Lewontin (1974) said, "to concentrate only on genetic change, without attempting to relate it to the kinds of physiological, morphogenetic, and behavioral evolution that are manifest in the fossil record, is to forget entirely what it is we are trying to explain in the first place". (p. 23).

We argue that in a similar vein, to concentrate only on the formulation of evolutionary theory of the MS in evolution education is to forget entirely what it is we are trying to achieve in evolution education and education in the broadest sense. In light of the successful applications of evolutionary concepts and methods across disciplines as diverse as economics, anthropology, history, psychology, and computer science, gene-centered evolution education is increasingly climbing the wrong mountain (*sensu* Hanisch & Eirodsh, 2020a). Given the urgency of evolving adaptive learning environments for all humans on the planet, we suggest evolution education specialists and interdisciplinary education innovators and policy makers work together to advance a new vision regarding the role of evolutionary concepts in the general education curriculum.

References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning Through Case Comparisons: A Meta-Analytic Review. Educational Psychologist, 48(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and Evaluation of the Conceptual Inventory of Natural Selection. Journal of Research in Science Teaching, 10(10), 952–978. https://doi.org/10.1002/tea.10053

Andrews, T. M., Kalinowski, S. T., & Leonard, M. J. (2011). "Are Humans Evolving?" A Classroom Discussion to Change Student Misconceptions Regarding Natural Selection. Evolution: Education and Outreach, 4(3), 456–466. https://doi.org/10.1007/s12052-011-0343-4

Apodaca, M. J., McInerney, J. D., Sala, O. E., Katinas, L., & Crisci, J. V. (2019). A Concept Map of Evolutionary Biology to Promote Meaningful Learning in Biology. The American Biology Teacher, 81(2), 79–87. https://doi.org/10.1525/abt.2019.81.2.79

Araújo, L. A. L. (2020). The centrality of evolution in biology teaching: towards a pluralistic perspective. Journal of Biological Education, 00(00), 1–12. https://doi.org/10.1080/00219266.2020.1757486

Baack, K., Becker, A., Eckebrecht, D., Kießling, J., Koch, M., Maier, A., & Roßnagel, G. (2016). Natura Oberstufe. Biologie für Gymnasien. Stuttgart: Ernst Klett Verlag.

Baedke, J., Fábregas-Tejeda, A., & Vergara-Silva, F. (2020). Does the extended evolutionary synthesis entail extended explanatory power? Biology and Philosophy, 35(1), 1–22. https://doi.org/10.1007/s10539-020-9736-5

Baraghith, K. (2020). Investigating populations in generalized Darwinism. Biology and Philosophy, 35(1). https://doi.org/10.1007/s10539-020-9735-6

Baraghith, K., & Feldbacher-Escamilla, C. J. (2021). The Many Faces of Generalizing the Theory of Evolution. American Philosophical Quarterly, 58(1), 35–50.

Barnes, M. E., Evans, E. M., Hazel, A., Brownell, S. E., & Nesse, R. M. (2017). Teleological reasoning, not acceptance of evolution, impacts students’ ability to learn natural selection. Evolution: Education and Outreach, 10(1), 7. https://doi.org/10.1186/s12052-017-0070-6

Booth Sweeney, L., & Sterman, J. D. (2007). Thinking about systems: student and teacher conceptions of natural and social systems. System Dynamics Review, 23(2/3), 285–311. https://doi.org/10.1002/sdr.366

Sweeney, L. B. (2006). How Is This Similar to That? The skill of recognizing parallel dynamic structures on center stage. The Creative Learning Exchange, 14(3).

Bray, J. H. (2010). Psychology as a Core Science, Technology, Engineering, and Mathematics (STEM) Discipline. Report of the American Psychological Association 2009 Presidential Task Force On the Future of Psychology as a STEM Discipline (Issue June).

Bruckermann, T., Fiedler, D., & Harms, U. (2020). Identifying precursory concepts in evolution during early childhood – a systematic literature review. Studies in Science Education, 00(00), 1–43. https://doi.org/10.1080/03057267.2020.1792678

Bruner, J. (1974). Toward a Theory of Instruction. Harvard University Press.

Centola, D., Wilensky, U., & McKenzie, E. (2000). A hands-on modeling approach to evolution: Learning about the evolution of cooperation and altruism through multi-agent modeling-The EACH Project. Proceedings of the Fourth International Conference of the Learning Sciences, 166–173.

Chesebrough, C., Davachi, L., Rock, D., Slaughter, M., & Grant, H. (2019). Coherence. The architecture of efficient learning. NeuroLeade, February.

Davies, J. (2014). A closed loop. The DNA helix gave 20th-century biology its symbol. But the more we learn, the more life circles back to an older image. Aeon. https://aeon.co/essays/the-feedback-loop-is-a-better-symbol-of-life-than-the-helix

Dweck, C. S. (2012). Mindsets and human nature: Promoting change in the Middle East, the schoolyard, the racial divide, and willpower. The American Psychologist, 67(8), 614–622. https://doi.org/10.1037/a0029783

Eirdosh, D., & Hanisch, S. (2021). Evolving Schools in a Post-pandemic Context. In W. Filho Leal (Ed.), COVID-19: Paving the Way for a More Sustainable World (pp. 465–480). Springer. https://doi.org/10.1007/978-3-030-69284-1_24

Erickson, H. L., Lanning, L. A., & French, R. (2017). Concept-Based Curriculum and Instruction for the Thinking Classroom (2nd ed.). Corwin.

Evans, E. M., & Rosengren, K. (2018). Cognitive Biases or Cognitive Bridges? Intuitive Reasoning in Biology. In K. Kampourakis & M. J. Reiss (Eds.), Teaching Biology in Schools. Global Research, Issues, and Trends. Routledge. https://doi.org/10.4324/9781315110158

Feist, G. J. (2006). The Psychology of Science and the Origins of the Scientific Mind. Yale University Press.

Feist, G. J. (2013). The psychology of scientific thought and behavior. The Psychologist, 26(12), 864–867.

Fuentes, A. (2021). “The descent of man,” 150 years on. Science, 372(6544), 769. https://doi.org/10.1126/science.abj4606

Gentner, D., Holyoak, K. J., & Kokinov, B. K. (2001). The Analogical Mind. Perspectives from Cognitive Science. MIT Press.

Gentner, D., & Maravilla, F. (2018). Analogical Reasoning. In L. J. Ball & V. A. Thompson (Eds.), International Handbook of Thinking & Reasoning (pp. 186–203). Psychology Press.

Gentner . D., & Colhoun, J. (2010.) Analogical Processes in Human Thinking and Learning. In Glatzeder, B., Goel, V., & Müller, A. (eds). Towards a Theory of Thinking Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03129-8

Gleick, J. (1987). Chaos. Making a new science. Viking.

Glynn, S. M. (2008). Making science concepts meaningful to students: Teaching with analogies. In S. Mikelskis-Seifert, U. Ringelband, & M. Brückmann (Eds.), Four Decades of Research in Science Education: From Curriculum Development to Quality Improvement (pp. 113–125). Waxmann.

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a Theory of Causality. Psychological Review, 118(1), 110–119. https://doi.org/10.1037/a0021336

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (1999). The scientist in the crib. HarperCollins.

Gopnik, A., Sobel, D. M., Danks, D., Glymour, C., Schulz, L. E., & Kushnir, T. (2004). A Theory of Causal Learning in Children: Causal Maps and Bayes Nets. Psychological Review, 111(1), 3–32. https://doi.org/10.1037/0033-295X.111.1.3

Gopnik, A., & Welllman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms and the theory theory. Psychological Bulletin, 138(6), 1085–1108. https://doi.org/10.1037/a0028044

Gregory, T. R. (2009). Understanding Natural Selection: Essential Concepts and Common Misconceptions. Evolution: Education and Outreach, 2(2), 156–175. https://doi.org/10.1007/s12052-009-0128-1

Gregory, T. R., Elliott, T. A., & Linquist, S. (2016). Why Genomics Needs Multi-Level Evolutionary Theory. In N. Eldredge, T. Pievan, E. Serrell, & I. Tëmkin (Eds.), Evolutionary Theory: A Hierarchical Perspective (pp. 137–150). University of Chicago Press.

Grotzer, T. A., Solis, S. L., Tutwiler, M. S., & Cuzzolino, M. P. (2017). A study of students' reasoning about probabilistic causality: Implications for understanding complex systems and for instructional design. Instructional Science, 45(1), 25–52. https://doi.org/10.1007/s11251-016-9389-6

Hanisch, S., & Eirdosh, D. (2020a). Educational potential of teaching evolution as an interdisciplinary science. Evolution: Education and Outreach, 13(1), 25. https://doi.org/10.1186/s12052-020-00138-4

Hanisch, S., & Eirdosh, D. (2020b). Conceptual clarification of evolution as an interdisciplinary science. EdArXiv Preprint. https://doi.org/10.35542/osf.io/vr4t5

Hanisch, S., & Eirdosh, D. (2020c). Causal Mapping as a Teaching Tool for Reflecting on Causation in Human Evolution. Science & Education. https://doi.org/10.1007/s11191-020-00157-z

Harrison, A. G. (2006). The Affective Dimension of Analogy. In Aubusson, P. J., Harrison, A. G., & Ritchie, S. M. (eds). Metaphor and Analogy in Science Education. Springer. https://doi.org/10.1007/1-4020-3830-5_1

Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. International Journal of Science Education, 22(9), 1011–1026. https://doi.org/10.1080/095006900416884

Harrison, A. G., & Treagust, D. F. (2006). Teaching and Learning with Analogies: Friend or foe? In Aubusson, P. J., Harrison, A. G., & Ritchie, S. M. (eds). Metaphor and Analogy in Science Education. Springer. https://doi.org/10.1007/1-4020-3830-5_1

Haskell, R. E. (2000). Transfer of Learning. Cognition, Instruction, and Reasoning. Academic Press.

Hayes, S. C., Barnes-Holmes, D., & Wilson, K. G. (2012). Creating a science more adequate to the challenge of the human condition. Journal of Contextual Behavioral Science, 1, 1–16. https://doi.org/10.1016/j.jcbs.2012.09.004

Hayes, S. C., Sanford, B. T., & Chin, F. T. (2017). Carrying the baton: Evolution science and a contextual behavioral analysis of language and cognition. Journal of Contextual Behavioral Science, 6(3), 314–328. https://doi.org/10.1016/j.jcbs.2017.01.002

Heddy, B. C., & Sinatra, G. M. (2013). Transforming misconceptions: Using transformative experience to promote positive affect and conceptual change in students learning about biological evolution. Science Education, 97(5), 723–744. https://doi.org/10.1002/sce.21072

Hofstadter, D. (2001). Analogy as the Core of Cognition. In Gentner, D., Holyoak, K. J., & Kokinov, B. K. (2001). The Analogical Mind. Perspectives from Cognitive Science. MIT Press.

Hoppitt, W., & Laland, K.N. (2013). Social Learning: An Introduction to Mechanisms, Methods, and Models. Princeton University Press.

Jacobson, M. J., & Wilensky, U. (2006). Complex Systems in Education: Scientific and Educational Importance and Implications for the Learning Sciences. Journal of the Learning Sciences, 15(1), 11–34. https://doi.org/10.1207/s15327809jls1501_4

Jablonka, E., & Lamb, M. J. (2005). Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life. MIT Press.

Jaenicke, J., & Paul, A. (2004). Biologie Heute entdecken S II. Braunschweig: Westermann Schroedel.

Kashdan, T. B., & Rottenberg, J. (2010). Psychological Flexibility as a Fundamental Aspect of Health. Clin Psychol Rev., 30(7), 865–878. https://doi.org/10.1016/j.cpr.2010.03.001

KMK (Kultusministerkonferenz) (2004). Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss.Luchterhand.

Lakoff, G., & Johnson, M. (1980). Metaphors We Live By. The University of Chicago Press.

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of Nature of Science Questionnaire: Toward Valid and Meaningful Assessment of Learners' Conceptions of Nature of Science. Journal of Research in Science Teaching, 39(6), 497–521. https://doi.org/10.1002/tea.10034

Legare, C. H., Opfer, J. E., Busch, J. T. A., & Shtulman, A. (2018). A Field Guide for Teaching Evolution in the Social Sciences. Evolution and Human Behavior, 39(3), 257–268. https://doi.org/10.1016/j.evolhumbehav.2018.01.002

Legare, C. H., & Shtulman, A. (2018). Explanatory pluralism across cultures and development. In J. Proust & M. Fortier (Eds.), Metacognitive Diversity: An Interdisciplinary Approach (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780198789710.003.0019

Lehrer, R., & Schauble, L. (2012). Seeding evolutionary thinking by engaging children in modeling its foundations. Science Education, 96(4), 701–724. https://doi.org/10.1002/sce.20475

Lewontin, R. C. (1974). The Genetic Basis of Evolutionary Change, Columbia biological series no. 25. New York: Columbia University Press.

Love, A. C. (2013). Interdisciplinary Lessons for the Teaching of Biology from the Practice of Evo-Devo. Science and Education, 22(2), 255–278. https://doi.org/10.1007/s11191-011-9351-6

Markl, J. (2018). Markl Biologie Oberstufe. Stuttgart: Ernst Klett Verlag.

Mead, R., Hejmadi, M., & Hurst, L. D. (2017). Teaching genetics prior to teaching evolution improves evolution understanding but not acceptance. PLoS Biology, 15(5), 1–30. https://doi.org/10.1371/journal.pbio.2002255

Mead, L. S., Kohn, C., Warwick, A., & Schwartz, K. (2019). Applying measurement standards to evolution education assessment instruments. Evolution: Education and Outreach, 12(1). https://doi.org/10.1186/s12052-019-0097-y

Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur (2014). Lehrpläne für die naturwissenschaftlichen Fächer für die weiterführenden Schulen in Rheinland-Pfalz. Biologie Chemie Physik. Klassenstufen 7 bis 9/10.

Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur (2016). Lehrplan für die Gesellschaftswissenschaftlichen Fächer. Erdkunde, Geschichte, Sozialkunde.

Moya, C., Boyd, R. T., & Henrich, J. (2015). Reasoning About Cultural and Genetic Transmission: Developmental and Cross-Cultural Evidence From Peru, Fiji, and the United States on How People Make Inferences About Trait Transmission. Topics in Cognitive Science, 7(4), 595–610. https://doi.org/10.1111/tops.12163

National Council for the Social Studies. (2013). The College, Career, and Civic Life (C3) Framework for Social Studies State Standards: Guidance for Enhancing the Rigor of K-12 Civics, Economics, Geography, and History. National Council for the Social Studies. https://www.socialstudies.org/sites/default/files/c3/c3-framework-for-social-studies-rev0617.pdf

National Research Council. (2012). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas (National Research Council (ed.)). National Academies Press. https://doi.org/10.17226/13165

Nehm, R. H., Beggrow, E. P., Opfer, J. E., & Ha, M. (2012). Reasoning About Natural Selection: Diagnosing Contextual Competency Using the ACORNS Instrument. The American Biology Teacher, 74(2), 92–98. https://doi.org/10.1525/abt.2012.74.2.6

Nelson, C. E., Scharmann, L. C., Beard, J., & Flammer, L. I. (2019). The nature of science as a foundation for fostering a better understanding of evolution. Evolution: Education and Outreach, 12(1), 1–16. https://doi.org/10.1186/s12052-019-0100-7

NGSS Lead States. (2013a). Next Generation Science Standards: For States, by States. National Academies Press.

NGSS Lead States. (2013b). Appendix G – Crosscutting Concepts.

NGSS Lead States. (2013c). Appendix J - Science, Technology, Society and the Environment.

Noble, D. (2006). The Music of Life. Biology beyond the Genome. Oxford University Press.

OECD. (2020). Curriculum Overload. A way forward. In Curriculum Overload. https://doi.org/10.1787/3081ceca-en

Oh, Y., Chesebrough, C., Erickson, B., Zhang, F., & Kounios, J. (2020). An insight-related neural reward signal. NeuroImage, 214(March), 116757. https://doi.org/10.1016/j.neuroimage.2020.116757

Pobiner, B. (2016). Accepting, understanding, teaching, and learning (human) evolution: Obstacles and opportunities. American Journal of Physical Anthropology, 159, S232–S274. https://doi.org/10.1002/ajpa.22910

Pobiner, B., Beardsley, P. M., Bertka, C. M., & Watson, W. A. (2018). Using human case studies to teach evolution in high school A. P. biology classrooms. Evolution: Education and Outreach, 11(3). https://doi.org/10.1186/s12052-018-0077-7

Pugh, K. J., Koskey, K. L. K., & Linnenbrink-Garcia, L. (2014). High school biology students' transfer of the concept of natural selection: a mixed-methods approach. Journal of Biological Education, 48(1), 23–33. https://doi.org/10.1080/00219266.2013.801873

Reydon, T. A. C. (2021). Misconceptions, conceptual pluralism, and conceptual toolkits: bringing the philosophy of science to the teaching of evolution. European Journal for Philosophy of Science, 11(2). https://doi.org/10.1007/s13194-021-00363-8

Roberts, N. (1978). Teaching Dynamic Feedback Systems Thinking: An Elementary View. Management Science, 24(8), 836–843. https://doi.org/10.1287/mnsc.24.8.836

Rosengren, K., Brem, S. K., Evans, E. M., & Sinatra, G. M. (2012). Evolution Challenges: Integrating Research and Practice in Teaching and Learning about Evolution. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199730421.001.0001

Russell, T., & McGuigan, L. (2019). Developmental Progression in Learning About Evolution in the 5–14 Age Range in England. In U. Harms & M. J. Reiss (Eds.), Evolution

Education Re-considered (pp. 59–80). Springer. https://doi.org/10.1007/978-3-030-14698-6_4

Sapolsky, R. M. (2018). Behave. The Biology of Human at Our Best and Worst. Vintage Books.

Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing Minds? Implications of Conceptual Change for Teaching and Learning about Biological Evolution. Evolution: Education and Outreach, 1(2), 189–195. https://doi.org/10.1007/s12052-008-0037-8

Smithsonian Institution (2015). What Does It Mean to Be Human? Student Workbook.

Stanley, S. (2020). Cultural Evolutionary Theory and the Significance of the Biology-Culture Analogy. Philosophy of the Social Sciences, October. https://doi.org/10.1177/0048393120971554

Stern, J., Ferraro, K., & Mohnkern, J. (2017). Tools for Teaching Conceptual Understanding, Secondary. Designing Lessons and Assessments for Deep Learning. Corwin.

Stern, J., Ferraro, K., Duncan, K., & Aleo, T. (2021). Learning That Transfers: Designing Curriculum for a Changing World. Corwin.

Tavoni, A., & Levin, S. (2014). Managing the climate commons at the nexus of ecology, behaviour and economics. Nature Climate Change, 4(12), 1057–1063. https://doi.org/10.1038/nclimate2375

Thagard, P., & Shelley, C. (2001). Emotional Analogies and Analogical Inference. In Gentner, D., Holyoak, K. J., & Kokinov, B. K. (eds). The Analogical Mind. Perspectives from Cognitive Science. MIT Press.

Thagard, P. (2007). Coherence, Truth, and the Development of Scientific Knowledge. Philosophy of Science, 74(January), 28–47.

Thanukos, A. (2008). A Look at Linguistic Evolution. Evolution: Education and Outreach, 1(3), 281–286. https://doi.org/10.1007/s12052-008-0058-3

University of Bath (2020). Understanding and Teaching Evolution. https://www.futurelearn.com/courses/understanding-evolution (accessed on 2 January 2022).

University of California Museum of Paleontology (2009a). Conceptual Framework. https://evolution.berkeley.edu/teach-evolution/conceptual-framework/

University of California Museum of Paleontology (2009b). A look at linguistic evolution. https://evolution.berkeley.edu/teach-resources/a-look-at-linguistic-evolution

Van Bouwel, J., & Weber, E. (2008). A pragmatist defense of non-relativistic explanatory pluralism in history and social science. History and Theory, 47(2), 168–182. https://doi.org/10.1111/j.1468-2303.2008.00445.x

Veblen, T. (1898). Why is economics not an evolutionary science? Quarterly Journal of Economics, 12(4), 373–397. https://doi.org/10.2307/1882952

Vendetti, M. S., Matlen, B. J., Richland, L. E., & Bunge, S. A. (2015). Analogical Reasoning in the Classroom: Insights From Cognitive Science. Mind, Brain, and Education, 9(2), 100–106. https://doi.org/10.1111/mbe.12080

Walker, C. M., Hubachek, S. Q., & Vendetti, M. S. (2018). Achieving abstraction: Generating far analogies promotes relational reasoning in children. Developmental Psychology, 54(10), 1833–1841. https://doi.org/10.1037/dev0000581

Zeidler, D. L., Herman, B. C., & Sadler, T. D. (2019). New directions in socioscientific issues research. Disciplinary and Interdisciplinary Science Education Research, 1(1), 1–9. https://doi.org/10.1186/s43031-019-0008-7

Ziadie, M. A., & Andrews, T. C. (2018). Moving evolution education forward: A systematic analysis of literature to identify gaps in collective knowledge for teaching. CBE Life Sciences Education, 17(1). https://doi.org/10.1187/cbe.17-08-0190

Part II: Generalizations of evolutionary theory: common principles or explanatory structures?

Chapter 9

From Games to Graphs. Evolving Networks in Cultural Evolution

Karim Baraghith[i]

Abstract

What is it that *evolves* in cultural evolution? This is a question easily posed but not so easily answered. According to common interpretations of cultural evolutionary theory, it is not strictly agents that change over time or proliferate during cultural transmission, but their socially transmitted behavior, what they communicate or acquire via social learning – in short: their *interactions.* This means that we have to put these cultural interactions into an evolutionary setting and show how they evolve within cultural populations, i.e. within social networks. But the social networks themselves also evolve, which brings us to a multi-level approach of cultural evolution, implementing both, individual and group selection. In this paper I will assume that the microlevel is given by a description of cultural agents, their behavior and decisions, whereas the macrolevel describes the dynamics on population structure and in particular population boundaries in social networks (since we are not really able to identify something analogous to 'species' in cultural evolution). In this paper, I am going to offer a specific mathematical model, that makes use of game theory for representing the cultural microlevel and graph theory for the cultural macrolevel. It has to be shown, how both can formally be linked in a synthetic attempt.

Keywords: cultural evolution; population; agents; graph theory; game theory

---

[i] Düsseldorf Center for Logic and Philosophy of Science (DCLPS), Department of Philosophy, Heinrich Heine University Düsseldorf.

9.1. Introduction

What is it that *evolves* in cultural evolution? This is a question easily posed but not so easily answered. According to common interpretations of cultural evolutionary theory, it is not strictly people or agents that change over time or proliferate during cultural transmission (Boyd & Richerson, 1985), (Mesoudi, 2011), (Lewens, 2015), but their socially transmitted behavior, what they communicate or acquire via social learning. In short: their *interactions.* This means that we have to put these cultural interactions into an evolutionary setting and show how they evolve within cultural populations, i.e. social networks. But the social networks themselves also evolve, which brings us to a multi-level approach of cultural evolution, implementing both, individual and group selection.

Borrowing from biology, we are inclined to adopt the common distinction between 'microevolution' and 'macroevolution'. This terminology however is problematic, since in evolutionary biology, changes in the population structure (which we will investigate in this paper) are universally viewed as an example of microevolution and not macroevolution. You could argue that the term 'macroevolution' in culture should be reserved for the emergence and diversification of cultural variation at or above the level of the society. However, there is no consensus in cultural evolutionary theory, how the taxonomic hierarchy is constituted are and where the macrolevel starts and the microlevel ends. All analogies have limits and should not be taken too seriously.

In this paper, we will speak of the cultural evolutionary 'microlevel' for mechanisms of cultural inheritance and 'macrolevel' as changes in population structure. I suggest to understand the first as the precise mechanisms of inheritance, while the second depicts the whole network and its evolving boundaries. That cultural evolution may have a micro- and a macrolevel has been stressed by many proponents of the approach, especially by (Mesoudi, 2011) or (Mace & Jordan, 2011).

In what follows, I am going to present some ideas for a model, which shall formally explain tools from the cultural macrolevel in terms of the cultural microlevel. In particular, I want to focus on some useful tools taken from *game theory* (micro) and *graph theory* (macro). My goal in this paper is to communicate ideas, not a precise mathematical proof for "games on graphs", which has been done elsewhere in different frameworks and contexts, see e.g. (Lieberman, et al., 2005), (Nowak, 2006) or (Débarre, et al., 2014). The idea here is to implement an evolving graph: unlike many models of evolutionary graph theory, here it is not the identities of nodes that change over times, but the presence/absence of edges between nodes and the type of interaction along these edges. In other words, the structure of the graph itself changes over time; this feature is described in the paper as macrolevel.

In section 2, I will briefly explore the microlevel via evolutionary game theory, its basic concepts, assumptions and predictions. Since different game theoretical strategies can (and have) been seen as different competing cultural variants and population dynamics and evolutionary game theory are therefore seamlessly transferable into each other (Hofbauer & Sigmund, 1998), the tools and concepts provided by the latter are highly valuable to describe the cultural evolutionary microlevel. In particular, I will focus on two specific games, that are quite common: the prisoner's dilemma and signaling games.

In section 3, I will turn the attention to the macrolevel, cultural populations and social networks. In biology "speciation" is what distinguishes micro- from macroevolution but I want to emphasize that due to a comparably high amount of so called "horizontal transmission" which we face in cultural evolution, species understood as reproductive communities cannot really be distinguished in cultural evolution. However, structural *populations* (and metapopulations) can. I will propose that populations of sociocultural systems should not be clustered by their type, but via their degree of causal connectivity. In this way, we can investigate population structure in social networks, to make links to the microlevel possible.

Finally, one such a possible link will be described in section 4, where I try to formally synthesize section 2 and 3. Explaining microlevel effects in terms of macolevel patterns and structures *synthesizes* evolutionary explanations. In order to accomplish it, one must determine and define what precisely it is that evolves on the cultural microlevel and how that has an influence on the microlevel and vice versa. The general idea which will guide us is that evolution will favor those groupings and populations, that yield a larger inner cohesion, i.e. the members are connected more firmly by their interactions.

9.2. The Cultural Microlevel: Agents playing Evolutionary Games

Cultural microevolution is about agents making decisions and proliferating their behavior socially. Game theory is a proper tool to model these kinds of social mechanisms since it identifies different kinds of social behavior as different (competing) *strategies*. Therefore, in this section, we will briefly introduce the rationale behind evolutionary game theory as a most promising toolbox for the study of cultural microevolution. I will provide two well-known examples from this research domain: the "Prisoner's Dilemma" and the "Signaling Game". Many other examples would have been possible, but to facilitate readability, we chose prominence over originality.

In general, game theory is a commonly used method applied throughout the social sciences, from economics to social psychology and political science.[1] *Evolutionary* game theory was famously introduced by (Maynard Smith, 1982) and investigated in detail e.g. by (Weibull, 1995). In general, game theory models the outcome ("payoff") of different mutual interactions ("strategies") between agents ("players"), that are rational in the sense of maximizing their "utility". A "solution concept" for a specific game thereby indicates how well a certain strategy will perform against another strategy. A well-known solution concept is the so called "Nash equilibrium", which is always the mutual "best response" to another players strategy.

The most important difference between standard game theory and evolutionary game theory is that the latter interprets the "utility" functions as fitness values and the rational considerations of the players are replaced by the optimizing force of selection on the long run. This basically implies two things: First, evolutionary game theory always concerns iterated games, that are played over many rounds (in the following, we will use the term: 'generation'). Second, the players do not need to have (conscious) prospects and preferences, they do not even have to know that they are playing a specific game, at all. In a more passive manner, they will still maximize their utility over time. Iterated Nash equilibria thereby become "evolutionary stable strategies".

Why is game theory such a promising candidate for the description of cultural microevolution? Because the competing behavioral variants that undergo a process of variation, selection and reproduction in cultural evolution – the "memes" if you are fond of this

---

[1] For a more detailed description of game theory in general, decision theory and its different areas of application in the (philosophy of the) social sciences, see (Steele, 2014).

terminology[2] – can in fact be interpreted as the different competing game theoretical strategies. Depending on the nature of the specific game under consideration, the theory will provide a dynamical analysis of the proliferation of these different variants/strategies. The idea that population dynamics and evolutionary game theory are seamlessly transferable into each other has e.g. been highlighted by (Hofbauer & Sigmund, 1998).

There are three major different kinds of evolutionary games, see (Weibull, 1995, pp. 28-31): a) defection/cooperation games, b) coordination games and c) hawk/dove (fight/compromise) games. In this paper, we will merely concentrate on the first two for the purpose of exemplifying. Let us now briefly lay out the basic framework of two very well-known games, before we interpret them in a *network style* in section 4.

9.2.1. The Prisoner's Dilemma and the Paradox of Altruism

There exist a multitude of different evolutionary games. One of the most famous games is the so called "Prisoner's Dilemma" (PD), a defection/cooperation game. The game is entirely symmetric, meaning that the "row" player (R) and the "column" player (C) have access to exactly the same repertoire of strategies, namely "cooperate" (altruistic) or "defect" (egoistic).

Table 1*: The famous "prisoner's dilemma" (PD),*
a symmetric game.

R / C	Cooperate	Defect
Cooperate	3/3	-1/4
Defect	4/-1	0/0

---

[2] For a more detailed analysis of the meme concept and cultural replicators, and arguments for its usefulness or irrelevance, see e.g. (Aunger, 2000), (Szathmáry & Fernando, 2010), (Henrich, et al., 2008), (Lewens, 2015) or (Dennett, 2018).

PD's payoff matrix (figure 1) states that pure cooperation pays off three times more than pure defection, but defection of one player given that the other cooperates, heavily pays off for the defector.[3] If both players are unaware of what the other one will do, it is always better to defect, even though both risk to get a payoff of zero. PD has just one Nash equilibrium, namely D/D, because it is the only option where both players cannot fare worse, should the other player change his strategy. No player can do better by unilaterally switching to a different strategy. If R defects, C best also defects because she risks getting -1 otherwise, and vice versa. Both could cooperate and get $\begin{pmatrix}3\\3\end{pmatrix}$, but there will always be a high risk that the other one defects, in which case the cooperative player fares worst. This is the dilemma, which gave the game its particular name. PD is used frequently in recent debates around the evolution (and the paradox) of altruism, see e.g. (Maynard Smith, 1974) or (Killingback & Doebeli, 2002).

9.2.2. Signaling Games and the Emergence of Conventional Meaning

Signaling games (SG) are a typical example of (evolutionary) coordination games (kind b, above). They model semantic interactions between a *sender* (S) and a *receiver* (R) and address the question of how meaningful communication and *conventional meaning* can emerge. This view describes meaning as serving a certain *function* of language, namely, to facilitate coordinated behavior. Public or conventional meaning is a consequence of pragmatic factors. SG tell an evolutionary story of how *language*, meaning and content can emerge out of simple environments, where none of them existed before and almost without any common knowledge assumptions or "we-intentions" of the players involved. Philosopher David Lewis seminal work "Conventions" (Lewis, 1969) can be seen as the starting point for SG in philosophy. Researchers like (Skyrms, 2010), (Huttegger, 2008), (Mühlenbernd & Franke, 2014), (Wagner, 2012), (Zollman, et al., 2013) or (Franke & Wagner, 2014) among others analyzed and modeled SG in *evolutionary* terms. The reader might inquire this literature to get a deeper understanding

---

[3] The numerical value of the payoffs are just exemplary, they can be altered, as long as their ordinal arrangement remains the same.

of the model and all its possible extensions – such as costly signals or conflict of interests – which we cannot provide here.

The very standard form of SG consists of a set of two world states $W = \{\sigma_1, \sigma_2\}$, two messages $M = \{m_1, m_2\}$ that S can communicate to R, and two acts $A = \{a_1, a_2\}$, by which R can respond to the received message. A pure S strategy is a function $s : W \rightarrow M$ from states to messages; a pure R strategy is a function $r : M \rightarrow A$ from messages to acts. There are 4 pure S and 4 pure R strategies and consequently 16 pure *strategy profiles.* It is important to note, that only S can observe σ_i . R has no information about it but only if R chooses the “right” corresponding a_i communication is successful and *both* sustain a payoff of 1, denoted as $U = \binom{1}{1}$. They have to cooperate and coordinate their behavior, which means we have to look at the *combination* of their strategies:

s_1: m_1 if σ_1, m_2 if σ_2

s_2: m_2 if σ_1, m_1 if σ_2

s_3: m_1 if σ_1, m_1 if σ_2

s_4: m_2 if σ_1, m_2 if σ_2

r_1: a_1 if m_1, a_2 if m_2

r_2: a_2 if m_1, a_1 if m_2

r_3: a_1 if m_1, a_1 if m_2

r_4: a_2 if m_1, a_2 if m_2

For the (combined) payoff matrix, see figure 2. There can be states of no communication, partial communication and perfect communication between S and R. Perfect communication is stable, partial communication is vulnerable to neutral drift and no communication is always unstable. (Lewis, 1969) called strict Nash equilibria in SG *signaling systems*. The *meaning* of the messages that constitute such signaling systems have a very high probability to become conventional. Evolution leads the population to converge to one of the two signaling systems given this set of initial conditions.

Table 2: The normal form of the combined strategies of the 2x2 Signaling Game.

	r1	**r2**	**r3**	**r4**
s1	1/1 (I)	0/0	½/½	½/½
s2	0/0	1/1 (II)	½/½	½/½
s3	½/½	½/½	½/½	½ /½
s4	½/½	½/½	½ /½	½ /½

In general, the dynamics are agnostic about the cognitive capacities of the agents. The sender and the receiver co-evolve and co-adapt and the signal's meaning emerges from the conventions that facilitate the coordination of both. Naturally, the structure of human languages is far more complex than it is captured in these simple SG described so far. But the model is simple and allows for many possible extensions, such as *conflict of interest* or *costly signals* (Zollman, et al., 2013)

9.3. The Cultural Macrolevel: Populations and Evolutionary Graphs

Let us now turn our attention the cultural macrolevel. What are the selective units here? Could it be a cultural counterpart to something like biological *species*? In biology, speciation is what separates microevolution form macroevolution (Ridley, 2004). Species are biological kinds and units of biological classification, i.e. units of macroevolution. In biology, the "biological species concept" is widely accepted but not unchallenged. It is going back to the work of Ernst Mayr, who defined species as groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups (Mayr, 1963). It is important to note, that there usually exist intrinsic reproductive barriers between members of one species and members of another species, which means that members of two species are not merely separated by ecological (external) boundaries. They mostly cannot even potentially interbreed, which members of different populations that became isolated, still are capable of. A species is a certain unique kind of organism in the entire biosphere, while a population is all of the conspecific members in one ecosystem or area. This is a general distinction between a species and a population, where the former is a part of the taxonomic hierarchy, while the latter is not. While biologists argue about the right method for species classification, philosophers mostly debate the ontological status of species, asking questions such as: Are species natural kinds or individuals? There is a long and ongoing debate in theoretical biology and philosophy of biology on the issue of the "right" species concept is, see (Ereshefsky, 2017). Biologist (Hey, 2001) offers about twenty practicable options. Since this paper is not about the metaphysical aspect of biological species, we will not deepen the matter here and restrict our view of biological species mostly on Mayr's biological species concept according to which species boundaries are given by intrinsic reproductive barriers.

9.3.1. Why are there no "Species" in Cultural Evolution?

The concept of a "species" is hard to establish in cultural evolution. That we may face populations but not really species in cultural evolution has already been stressed out by (Boyd, et al., 1997). Although cultural traits may evolve through a process of variation, selection and reproduction and thereby resulting in fissions of cultural lineages that can be observed and structured, branches of the "tree of culture" can additionally *reunify* later on and frequently do so, for instance in the case of customs or traditions. Although there are geographical, linguistic or politic separations or specializations of cultural lineages, as soon as they get in contact again for one reason or another, there is a high probability for the initiation of a transmission *between* already established separated branches: ways of cultural interchange or cultural diffusion.

This "horizontal transmission" or diffusion also occurs in biological evolution. The important difference is the mere *amount* of diffusion in cultural evolution, that makes the application of phylogenetic methods very hard to establish. This is so, despite some "transmission isolating mechanisms" (TRIMS) between different linages, which we can sometimes find in culture. Out of a range of exemplar cases from the literature, let me pick one in particular: An empirical study by (Tehrani & Collard, 2013) indeed revealed that the knowledge of carpeting and textile production in Iranian tribal groups, who still live in a very traditional way, is mostly passed on within the group (which is mostly a large and ramified family) and kept there. Their study revealed TRIMS, which prevent cultural variants from diffusing horizontally, like e.g. the mothers teaching only their daughters the specific skills and carefully prevent their sons from learning it, since they often leave the family or group. But what if such TRIMS are absent or very weak such as in our globalized internet culture? In these (most relevant) cases of cultural evolution, we may need a very different more *network based* formal framework, largely implementing so called “small-world phenomena” (Kleinberg, 2004), which are in this context equivalent to horizontal transmission. The graph will differ from evolutionary trees, where information mostly just flows in one direction, from parents to offspring. That means we have to know more about the actual population structure.

If we take a closer look and allow for such a constant flow between lineages (horizontal transmission), the standard rules of cladistic classification break down, simply because we cannot determine a *common ancestor* (CA) of members of two different species anymore. This means that the important distinction between *homology* and *homoplasy* also collapses. The question arises, whether a macrolevel taxonomy is still diagnostically efficient, if this problem *frequently* occurs (let’s assume in slightly over 50% of cases of cultural transmission).

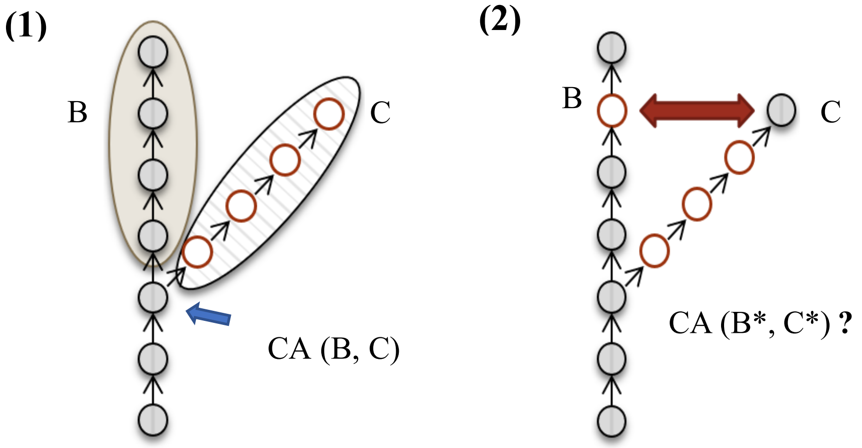


Figure 1*: The difference between (1) a classical biological tree of descent and (2) a cultural transmission graph. In (1), “species” (ellipses) can be separated after the event of “speciation”, that prevents interbreeding between individuals of the separated lineages, and constitutes a reproductive barrier. If such a barrier is generally missing, as in (2), and horizontal transmission (red arrow) frequently occurs, species cannot really be distinguished anymore and a common ancestor (CA) cannot be identified. Modified after (Schurz, 2011, p. 224).*

Let us take a look at figure 3. In case (1) species B and C clearly share a common ancestor (CA). But how is the situation in the almost analogue case (2)? Because of the horizontal flow of cultural information (red arrow), we have to assume B* and C* as intermediate ancestors.

B* is a partial ancestor, which cannot explain all the differences between B and C and the same is true for C*. The important distinction between homologies and homoplasies cannot be maintained here. If the side branch B* <–> C* injected a trait t* from B to C or vice versa, would t* then be a homology or a homoplasy? It may have evolved only in one branch of the tree (homology), but shortly after that, it will be found on a distinct branch, though it was not there before (homoplasy). Thus, the distinction collapses. If we cannot successfully demarcate groupings in the way we demarcate biological species in cultural networks, it seems that we are in need of another kind of clustering tool.

9.3.2. Interactive Groupings in Cultural Population Structure

What other concept could we establish in order to demarcate groupings in cultural evolution? I think that the key lies in successfully delimiting boundaries in cultural populations, i.e. social networks. According to Roberta Millstein's definition of a (biological) population, the "*boundaries of a population are those groupings where the rates of interactions are much higher within than without"* (Millstein, 2010, p. 67). She calls that the "causal interactionist population concept" or short the CIPC. In (Baraghith, 2020), it has been suggested to take this "inner interactive connectivity", i.e. *cohesion* in cultural population structure as the population defining criterion. In this reading, populations are a specific kind of "nearly decomposable systems", just in the sense of (Simon, 2002). The rates of interactions are lower between populations than within them and precisely this feature gives them their (sometimes fuzzy) boundaries. Among other things, this is precisely what makes populations *individuals*, that can be distinguished from other individuals, a point that has also been stressed out by (DiFrisco, 2018, p. 20). Near decomposability is a very powerful explanatory tool which can help to discover communities in complex networks of interactions, where in principle any agent can be connected with any other one. It is a great tool for "community discovery", as the problem of finding proper groupings in complex networks is e.g. called in the machine learning literature (Rossetti, et al., 2017).

Furthermore, it has been shown that under the usual conditions of mutation and natural selection, ND systems will increase in fitness, and therefore reproduce at a faster rate than systems that do not possess the property. This result has been proven mathematically on several occasions, e.g. (Simon, 2002, p. 588). Instead of repeating the proof here, let me just paraphrase it via an example. The general idea is that you compare two complex dynamical systems, one with an ND architecture and one without. Simon choses the example of two ways of creating a watch, see (Simon, 2002, p. 591). Two watchmakers want to create as many identical watches as possible, which consist of 1,000 parts each. However, the watches of the first watchmaker are composed of “10 stable subassemblies of 10 parts each” (ibid.), while the second set does not have this ND property. “Stability” in this sense can also be interpreted as having more internal interactions of any kind. If we now assume frequent and regular interruptions of the manufacture/production, that are more or less disruptive to both systems, it is easy to see that the second watchmaker will have to start again from scratch after every disruption, and will produce far less watches on the long run; he has to assemble 1,000 parts without disturbance. The first one however only must assemble 10 parts between the disruptions, he will almost never have to start again from scratch and will produce more watches. That is, on the long run, the ND systems will be much fitter and less prone.

Population structure is defined (in biology) by the topological organization of genetic variation and is driven by the combined effects of evolutionary processes that include recombination, mutation, genetic drift, demographic history, and natural selection. It shows how population boundaries within species are delineated, and the contribution of evolutionary processes to the formation of distinct (sometimes fuzzy) populations. The concept (or models that implement it) are hardly discussed in debates around cultural evolution, for an exception, see (Scapoli, et al., 2005). The structure of a population however, i.e. how exactly the network looks like, who interacts with whom etc. will most likely have an effect on which traits evolve and how. Population structure – when combined with population dynamics – can yield very interesting and informative insights that are hidden in standard population dynamics, for instance in the evolution of languages, see (Lee, et al., 2005). Another example is the well-known “paradox of altruism”: In structured populations, social behavior evolves if, for social individuals, the net social benefit of living next to other social individuals outweighs the costs of competing against them. This is the result of a graph theoretical investigation by (Débarre, et al., 2014).

That graph theory is a proper tool for the study of generalized evolutionary systems is shown in (Jagers op Akkerhuis, 2016), (Lieberman, et al., 2005) or (Nowak, 2006, p. 123). That groupings in complex networks can yield information about *biological* hierarchy has famously been shown by (Tëmkin & Eldredge, 2015). But we hardly find any literature on this quite new field of research within the framework of cultural evolution. In general, graph theory provides us with the opportunity to investigate the *topological features* of a network structure. Formally, any graph G is determined by the set of its nodes/elements V and the edges/connections E between nodes. We write: $G = (V, E)$. Let us take Millstein's nearly decomposable CIPC seriously, see (Millstein, 2010), and assume a population to be a set of agents $V = (v_1, v_2, \dots, v_n)$. The agents of the population are represented by the nodes of a graph, whereas the interactions between them are represented by the edges $E = (e_1, e_2, \dots, e_n)$. We could then think of defining an *index* for the population's density: a "cohesion index" CI_P for a population P, as (Baraghith, 2020) has done. We start with the simple idea:

$$CI_{abs}(P) = \frac{\#E_P}{\#V_P}$$

with $\#E_P$ indicating the number of actual edges in the population P and $\#V_P$ the number of nodes in P. $CI_{abs}(P)$ is the absolute interactions density of P. It is a *relation of edges and nodes* within P. $\#V_P$ is the number of the members of the population. The total maximal number of their possible interactions $(\#E_P)$ is clearly given mathematically by $\#V_P \times \frac{\#V_P-1}{2}$ and the maximal relation $max\ CI_{abs}(P)$ by $\frac{\#V_P \times \frac{\#V_P-1}{2}}{\#V_P} = \frac{\#V_P-1}{2}$.

$\#V_P$	$\max\#E_P$	$max\ CI_{abs}(P)$
1	0	0
2	1	$\frac{1}{2}$
4	6	$\frac{6}{4}$
5	10	$\frac{10}{5}$
10	45	$\frac{45}{10}$

The adjoining table lists some numerical examples. Now, a proper measure for CI_P could be to interpret it as a percentage of the maximum relation of $\#V_P$ and $\#E_P$. Therefore, we define:

$$CI(P) = \frac{CI_{abs}(P)}{max\ CI_{abs}(P)} = \frac{\frac{\#E_P}{\#V_P}}{\frac{\#V_{P-1}}{2}}$$

$CI(P)$ is the *relative density* of P, relative in terms of a percentage of the maximum density of possible interactions. We could determine a threshold, say $t = 60\%$. Every subgraph of P with $CI(P) \geq t$ is a proper population candidate in the sense of Millstein's CIPC. Summary definition*: A subgraph P of a graph G is a candidate for a population if P is maximal in the sense, that its relative interactions density is greater or equal t.*

9.4. Synthesis: Games on Graphs

Studying population dynamics within structured population is a crucial task. To quote Harvard biologist and mathematician Martin Nowak (Nowak, 2006, p. 124):

> The general question of how population structure affects evolutionary dynamics is hugely important and has been a longstanding topic in population genetics. A deeper, mathematical understanding of cultural evolution in human society requires the *study of evolutionary dynamics on social networks*.

The cohesion index CI_P which we developed at the end of the last section is able to reveal groups or communities in every possible cultural network, where cultural information (unlike in biological trees) potentially flows between *all* possible members of the network and no "species" boundaries can be determined. In this final section, we will try to synthesize our structural population graphs determined by CI_P with game theoretical (i.e. population dynamical) tools described earlier. This is a crucial requirement for any evolutionary explanation, (biological or cultural) since up to this point, our formalism CI_P does only provide a static, not a dynamical picture. We will include the prisoner's dilemma and the signaling game.

Let $G = (V, E)$ be a population graph at time t_1. We can introduce *reproduction, variation* and *selection* by comparing it to another graph $G' = (V', E')$ at time t_2 which is a result of G's change in time. For reasons of simplicity, we only assume a slight change in the topology. The nodes $V = (v_1, v_2, \dots, v_n)$ stay the same[4] while only the edges $E = (e_1, e_2, \dots, e_n)$ can disappear, i.e. become negatively selected (selection). Also, new edges emerge (variation). All other edges reappear in G' (reproduction). This idea differs from "standard" evolutionary graph theory, where variation, selection and reproduction are interpreted otherwise, see e.g. (Nowak, 2006), (Débarre, et al., 2014) or (Allen, et al., 2017).

The state space of *all possible states* that the dynamical system can take is a large set of possible combinations within the maximal number of edges $\#E_{max}$. The latter can be calculated

---

[4] One could also think of introducing variation, selection and reproduction of the nodes, not only the edges. This would mean that *new individuals* can appear or disappear in the populations under investigation. If one would interpret this appearance or disappearance as birth and death of the individuals, the model would indeed allow for a conceptual bridge between biological and cultural evolution. Future work will most probably consider these options and in some evolution-based *community discovery* algorithms such as TILES (Rossetti, et al., 2017), the idea of changing edges and nodes is already implemented. In this paper however, we will *restrict* the application to the evolution of the edges, i.e. interactions, for reasons of vividness.

by the total number of nodes $\#V$ via $\#E_{max} = \#V \times \frac{\#V-1}{2}$. All of these possibilities can be reached by the dynamical system, at least in principle, but there will be attractors. These are precisely those graphs and subgraphs, that yield a *higher inner cohesion*, i.e. have a high $CI(P)$. In other words: the larger its inner cohesion, the fitter the population on the long run.

For t, we assume a discrete sequence, i.e. countable reproductive cycles. The *state* of the system in this particular case is G's *topology* at time t. The evolutionary system parameters (variation, selection and reproduction), involve the relative 'fitness value'[5] of each possible graph G_i. Because we compare different possible populations/groups, we have a case of group selection. Furthermore, we will introduce two weighting functions, one for the players (nodes) and one for their interactions (edges). The overall idea is that *evolution will positively select those groupings, that have the larger cohesion*.

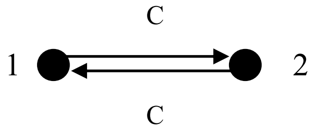
9.4.1. The Prisoner's Network Game (PN)

Let us first demonstrate the general idea on the famous Prisoner's dilemma (PD). Each player is represented by a node, each played strategy by an edge. Since game theory is always interactive, we have *diagraphs* now, i.e. there are *two* edges between two nodes, one representing what player 1 does, and one for the strategy of player 2. Furthermore, we change from undirected graphs (that were sufficient to demarcate populations so far) to *directed graphs*, where the flow of information is directional. We assume that both play at the same time. Consequently, we have the following options for PD:

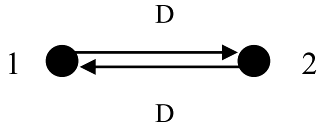
---

[5] A conceptual clarification: In biology, the *relative fitness* determines changes in genotype frequencies. It is given as $p(t+1) = \frac{\omega}{\bar{\omega}} p(t)$, where ω stands for fitness. The more often a variant dynamically occurs in a population, the higher its ω. In this paper, however, we understand 'fitness' of a graph in a different sense. We do not treat the graphs themselves as relevant variants, but the edges. The higher their frequency, the higher the total graphs 'internal fitness'.

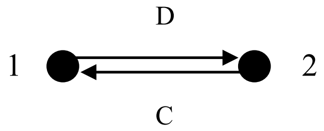
Both cooperate, the payoff is $\begin{pmatrix}3\\3\end{pmatrix}$.



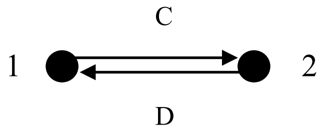
Both defect, the payoff is $\begin{pmatrix}0\\0\end{pmatrix}$ (only Nash-equilibrium).



Player 1 defects while player 2 cooperates: $\begin{pmatrix}4\\-1\end{pmatrix}$



Player 1 cooperates, while player 2 defects: $\begin{pmatrix}-1\\4\end{pmatrix}$



Now, let us put these options into a larger setting. Assume a social situation of any kind, e.g. a small company or a peer group. Five people are involved in it, see figure 4. We assume a random distribution of edges and nodes at the beginning. For reasons of simplicity, let us just observe this small graph at t_1. We assume the reproductive cycles (generations) of the strategies to be sufficiently long enough for a player V to play different strategies with different

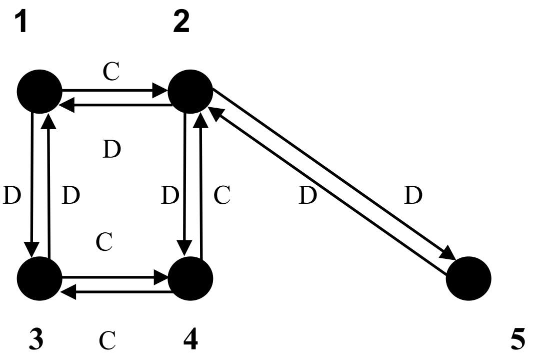


***Figure 2**: A small network of five players, each has the possibility to cooperate or to defect.*

other players. We therefore face the (quite realistic) case of *overlapping* generations, which are nevertheless discrete from the perspective of each single strategy.[6]

We can see that different strategies exist next to each other at the same time in the population. A population dynamical analysis would tell us that there are 6 D- and 4 C- strategies involved, therefore (if we assume a previous generation and error-free reproduction of the strategies) the *fitness* – simply given as the number of instantiations of a specific strategy – of D equals 6, and the fitness of C equals 4. Even though we just have a static picture so far, we can assume from the structure of the game itself, that the cooperators will die out more quickly, and D is on its way to fixation in the next generations (since the cooperators lose more utility while the defectors gain). So far, we are quite in line with classical findings of PD in evolutionary game theory.

But given figure 4, we know now a lot more about the situation. We know that player 3 interacts with player 4 and player 1, and that player 5 only interacts with one other player, while the other four interact with more. We know that player 2 is at the center of the population – she is the only one with $\deg 3$ – even though she only defects and never cooperates. These (possibly crucial) informations are hidden in standard population dynamics and this is probably one of the reasons, why cultural evolution and game theory were not able to explain a lot of fine grained social structures and phenomena so far, as adversive critiques of cultural evolutionary theory have often claimed, see e.g. (Lewontin, 2005).

Each strategy type (C, D) could now be classified not only by its average payoff, but more fine-grained, namely as *interaction-tokens* in relation to the specific player, who plays it. The number of each players options corresponds to his deg, i.e. how well she is connected, whereas the number of possible options for a player is given by all other nodes in the graph, *except her own* (one cannot socially interact with oneself). How much payoff she actually gets also depends on what strategies she chooses, of course. The partial payoffs have to be combined. For example: Player 4 cooperates two times, once with player 3 (who also cooperates) and once

---

[6] As (Weibull, 1995, p. 124) shows, generation overlapping can be modeled via the introduction of an effective generation rate $\tau > 1$, where only a τ fraction of the total population reproduces.

with player 2 (who defects). Therefore, player 4 will receive a total payoff of: $3 + (-1) = 2$ in the next generation.

For any player (i.e. node) V we write:

$$u_c(V) = \sum_{1\leq n}^{n} \left(u_1(V), u_2(V), \ldots, u_n(V)\right)$$

with n indicating a player's connective degree. Her total/cumulative utility u_c in one round equals the sum of the partial payoffs of her (overlapping) interactions e_n . Center player 2 fares best, she receives $u_c(2) = 4 + 4 + 0 = 8$.

It is precisely this sum, that introduces our *first weighting function* for the dynamical population. It shows us how good a player fares in one round, i.e. how much she benefits from her population structure and her own interactions. The more total utility a player gets, the higher her chances of keeping her strategy in the next round. This first weighting in the population structure is straightforward. Let us call it the "*player's weighting" (PW*), since it indicates how good a specific player fares in a round/generation. Naturally, if we involve more rounds, $T = (t_1, t_2, \ldots t_m)$, it will add up to

$$u_c^{t_c}(V) = \left(u_c^{t_1}(V) + u_c^{t_2}(V) + \cdots + u_c^{t_m}(V)\right),$$

with m indicating the number of rounds/generations and $u_c^{t_c}$ indicating the cumulative payoff of a player V in all rounds under consideration. Given the specific structure of the payoff matrix, this value can become negative, as well.

The second weighting is a bit more complex. In order to get the second weighting function, we have to introduce a dynamical analysis, i.e. we have to look at the next rounds. We take the distribution of figure 4 as t_1 and allow for changes in the players strategies. The players (whose number we will keep constant) will only keep their strategies in t_2 , if there proved well in the last round; in this sense, our players are rational. Otherwise they *stop playing* at all. This “do nothing” strategy, which we want to establish here is quite important, because it allows us to implement the cessation of an interactive *channel* between two players and therefore the Darwinian module of *selection*. Doing nothing is not as good as doing the right thing, but it is better than acting wrong. If, say, player 4 cooperated with player 2, but was defected by him $\begin{pmatrix} -1 \\ 4 \end{pmatrix}$, she will most likely stop this interaction after that. In order to formalize this “do nothing” strategy, our payoff matrix must be updated, see figure 5. We take the logical symbol for negation “¬”, to denote this new strategy.

***Table 3**: The “prisoner’s network game“ (PN), as I call it. A new strategy is introduced, namely ¬ (do nothing).*

	Cooperate	Defect	¬
Cooperate	3/3	-1/4	0/1
Defect	4/-1	0/0	0/1
¬	1/0	1/0	1/1

If ¬ meets another strategy, we assume a payoff of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. The player who defects or cooperates against ¬ , gets nothing, because the other player simply does not react at all. He wastes time or energy. ¬ on the other hand, receives 1 because she saves this time/energy. Accordingly, if both players play ¬ , the payoff is $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, because both save time and energy for possible interactions with other players in the network. Since the matrix of figure 5 does not show the standard PD anymore, we name this new game: *prisoner’s network game* (PN).

It is easy to see that $\neg/\neg$ constitutes a second Nash-equilibrium, next to D/D. Just like D/D however, it is not Pareto optimal, since both players would each fare better, if they played C/C. But instead of just defecting, there is now an incentive to do nothing instead. If we assume that each player reacts with $\neg$ to a *loss* that he received in the previous round, the network of the next round would change according to figure 6.

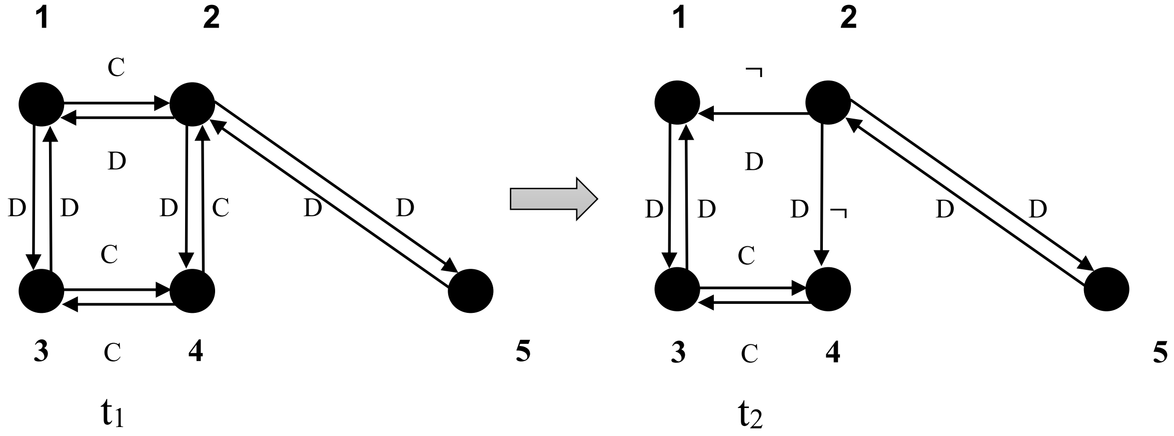


***Figure 3**: The second round of the PN game. Since interaction have to be mutual, an interaction channel ceases to exist if one side plays* $\neg$ *. Therefore, the graph decomposes into two independent graphs. Defection leads to decomposition.*

The graph has been decomposed into two independent graphs $G_{a2} = \big((1,3,4), E_a\big)$ and $G_{b2} = \big((2,5), E_b\big)$. Now it is not a full diagraph anymore since interactions have to be mutual and $\neg$ is not an interaction, even though it is a strategy. Also, the payoffs have changed dramatically. Player 2, that had the highest u_t at t_1 , now only receives $u_t = 0$ and only mutually interacts with 5 anymore, because both still exploit one another (D/D). Given by the payoff matrix, there is a high incentive now for player 2 to stop interacting with player 1 and player 4 in the next round t_3 .

Intermediate results: The one-round scenario (figure 4) revealed that defectors fare best (just as in the classical PD), and additionally, $u_c(V_i)$ depends on $deg\ (V_i)$. The more players to exploit, the better. However, under iteration the outcomes change. Players who defected a lot are now mostly being ignored, namely by the “betrayed” cooperators.

This result is similar, but not equivalent to results of the *iterated prisoner's dilemma,* that also differs from standard PD, see (Dal Bó & Fréchette, 2019).[7] For that, at least a small amount of *memory* of the players is required. However, scenarios of the iterated PD do not take population structure into account, which our account does. This development has severe consequences for the evolutionary macrolevel. It seems quite intuitive, that a high rate of cooperative interactions strengthens a population internally, while defection leads to decomposition. This adds a new possible solution to the *paradox of altruism*, apart from others, that have been discussed in the evolutionary literature, so far. In a nutshell: *dynamical population structure facilitates cooperation.*

Now, what could constitute the second kind of weighting (next to PW), that we spoke of? It is given by the *rate of interactions in a population, that get reproduced in the next round*. We shall call it the "*reproduced-interaction weighting" (RIW*). The more often a specific interaction (by which we mean mutual interaction) gets reproduced, the higher its specific RIW. An interaction becomes selected, if it ceases to exist in the next round, i.e. if one player plays $\neg$. For example, the mutual D interaction between player 1 and player 2 (figure 6) has RIW = 2 in the second round t_2 since it persisted for 2 rounds. RIW is therefore strictly proportional to the rounds that the network game is played already but once a strategy pair becomes negatively selected, RIW ceases to increase.

---

[7] Most prominent was Axelrod's strategy tournament in the 1980's, see (Axelrod, 1984), where – among many other sophisticated strategies – the "Tit-for-Tat" (TFT) meta-strategy usually sticked out, which "rewards" C with C and "punishes" D with D. Similar in our PN, just that the cooperators will not defect, if they were themselves defected, but *stop* interacting at all. I assume that this is in fact a more realistic scenario, but this is of course an empirical question.

So far, we implemented *reproduction* and *selection*. In order to implement the Darwinian module of *variation*, we have to allow for *new* interactions to appear in the next round. With a certain probability, a new interaction appears between any two players, that have not interacted in the previous round, where the total number of possible interactions for a specific player is given by the total number of nodes in the graph minus one (a player cannot interact with herself). If variation is assumed as random, a new interaction could be D, C, or ¬ , with equal probability. If however, more realistically, variation is assumed as *guided*,[8] it may very well be that the players *learn* from their past payoff and chose their new strategies hereof. In this case, players that mutually cooperated will most likely tend to cooperate again, while players who defected will most likely stop doing that after some number of rounds (since they get a bad payoff after a starting bonus in the first round and are mostly being ignored).

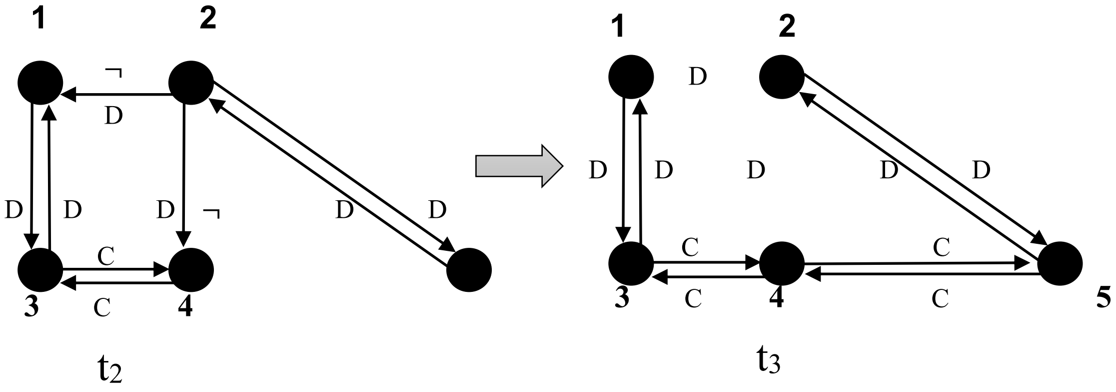


Figure 4*: The third round of the PN game. A new interaction between player 4 and player 5 has been established (variation). Since player 4 (a cooperator) received a stable outcome in the last two rounds, it is very likely that she opens up a new junction and since player 5 received nothing from defection so far, he reacts with cooperation, as well. In this way, cooperative behavior has the opportunity to spread in the population and will lead to more cohesive graphs on the long run.*

---

[8] In cultural evolution, guided variation is most common. Guided variation means that cultural variants are changed, before transmission, see (Boyd & Richerson, 1985, p. 81) or (Mesoudi, forthc.). In our interpretation, it can also mean that new variants can arise. Guided variation reflects a cultural agent's personal disposition, to change the cultural information that she receives or even create new information. This altering often is driven by *personal* experience, desires and beliefs.

In figure 7, we depict the next (third) round of our exemplar scenario. A new interaction between player 4 and player 5 has been established (variation). Since player 4 (a cooperator) received a stable outcome in the last two rounds, it is very likely that she opens up a new junction, and since player 5 received nothing from defection so far, she reacts with cooperation, as well. In this way, cooperative behavior has the opportunity to spread in the population and will lead to more cohesive graphs on the long run.

9.4.2. The Signaling Network Game (SN)

Let us now turn our attention to *signaling games*. In the classical Lewisian SG, the problem of the game was not of "egoistic vs altruistic" nature, but a problem of *coordination*. Sender (S) and receiver (R) have to coordinate their behavior to receive a stable payoff. For reasons of space, let us not list all 16 strategy combinations here, they could be depicted just like in the previous section. If S and R do not coordinate, they receive either nothing $\begin{pmatrix}0\\0\end{pmatrix}$, or mere coincidence decides $\begin{pmatrix}1/2\\1/2\end{pmatrix}$ (pooling equilibria). In both cases, no stable communication occurs and no stable language community can possibly evolve. In the two cases of perfect communication, however, a language will evolve, therefore these are the two Nash equilibria ("signaling systems"). Under iteration these become *conventions of meaning*. Evolution will lead the dynamical system to fixation of either the first *or* the second; which route will be taken depends on small coincidental fluctuations at the beginning and cannot be foreseen in advance. It is also not that important. Important for stable communication is just the fact, that S and R *do* coordinate and play reflexively symmetric. In figure 8, the normal form is depicted but note that we already implemented the "do nothing" strategy $\neg$. Therefore, the normal form does not depict the classical SG anymore and (in analogy to the previous section) we will call this new game: "signaling network game" (SN).

***Table 4**: The Signaling Network Game (SN). The blueprint is the standard 2x2 Lewis Signaling game with its two Nash equilibria I and II. Additionally, “do nothing” ¬ is implemented as a new strategy, and (III) constitutes a third Nash equilibrium.*

	r₁	**r₂**	**r₃**	**r₄**	¬
S1	1/1 (I)	0/0	½/½	½/½	0/¼
S2	0/0	1/1 (II)	½/½	½/½	0/¼
S3	½/½	½/½	½/½	½ /½	0/¼
S4	½/½	½/½	½ /½	½ /½	0/¼
¬	¼/0	¼ /0	¼/0	¼/0	¼/¼ (III)

Like in PN, ¬ grants $\begin{pmatrix} 0 \\ 1/4 \end{pmatrix}$ payoff for the player that faces it, but leads to $\begin{pmatrix} 1/4 \\ 1/4 \end{pmatrix}$ if both players play ¬ , because both save energy. Note that we slightly altered the payoffs numerically in contrast to the previous section, but the ordinal scale remains the same. Doing nothing is not the best strategy, but better than acting wrong.

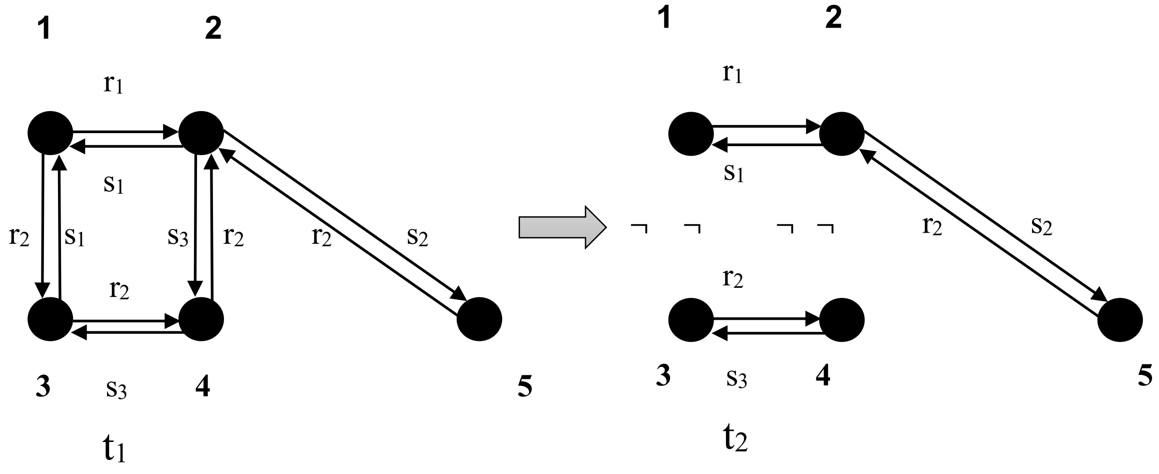


***Figure 5**: Two rounds of the SN game: players playing signaling systems will strictly prevail, while pooling equilibira will only be reproduced in 50% of the cases. All other interactions become negatively selected, since the players do not receive any payoff. An exeption is the mutual interaction between player 3 and player 4, which prevails, since it constitutes a pooling equilibrium. On the long run however, it will most likely dissapear.*

Let us now implement SN into our graph structure and let us take the same small graph as before. We will directly take a look at generations t_1 and t_2 . Some aspects are important now: A node will represent both, S and R, since an agent likewise sends *and* receives information during one round. Since (as before) interactions have to be mutual, one directed edge will always indicate a sender strategy (s_i), the other one a receiver strategy (r_i). There could be “pure senders” (nodes that only play s_i to all nodes that they are connected with) and “pure receivers” (only playing r_i) and “mixed players” (playing both kinds of strategies). We assume most players to execute mixed strategies. Figure 9 only depicts one possible option (out of many scenarios), but we can still deduce some rules for the dynamical graph.

In t_1 , the strategies are mixed randomly. But already in t_1 , one can observe that signaling systems (I and II above) become reproduced, while other strategies do not. The interaction between player 1 and player 3 yielded $\begin{pmatrix}0\\0\end{pmatrix}$, therefore, both players cease to interact. The interaction between player 1 and player 2 on the other hand was quite successful, such as between player 2 and player 5: $\begin{pmatrix}1\\1\end{pmatrix}$. Therefore, these interactions will prevail. What about all the other interactions, the so-called pooling equilibria? Since success in these interactions is random $\begin{pmatrix}{}^1/_2\\{}^1/_2\end{pmatrix}$, we also assume a 50% chance of such an interaction to be reproduced. Luck decides, whether the players get a payoff. One can see that like in the previous section, the graph has been decomposed in two independent subgraphs. In a coordination game like SN, successful communication will increase cohesion, while non-understanding destroys it. There will be an incentive (selective pressure) towards communication and therefore, players will most likely try to reconnect in the following rounds (which we will not depict here, the reader can imagine what is meant), until they also reached stable communication.

9.4.3. Dynamical Analysis for Cultural Population Structure

After observing two exemplar cases – the prisoner’s network game and the signaling network game – we can finally turn our focus on formulating the specific axioms for our model of population games on evolutionary graphs.

What effects do both weightings (PW and RIW) yield on the population structure? Remember (section 3) that the population boundaries are given by a higher internal interaction rate and a lower external interaction rate. The more internal interactions exist, the more cohesive the population, and the more clearly and unambiguously it can be identified in contrast to other populations. On the other hand: the more players in a population play anything else than $\neg$, the more interactions there are, since $\neg$ stops interactions and weakens the populations cohesion. Selection *favors* populations with a high CI(P). The more players play in a population, the higher its CI(P). And this factor (the number of players that actually interact) is given by PW and RIW, the higher their contribution to the dynamical system, the better. But how exactly does this dynamical system looks like in our context?

So far, we have described how cultural population graphs look like and how game theoretical features could be implemented in them. But the picture is not complete, yet. The following pages will indicate all ingredients/variables for a model of dynamical evolutionary graphs/populations (macrolevel) that contain evolutionary games (microlevel). As indicated in the table (figure 10), for the evolution of games on graphs, we implement the three Darwinian modules.

Reproduction r of a graph G is given as (already existing) edges, which manage to survive into the next generation. Such interactions are weighted by RIW, i.e.:

$$r(G_i)_{t+1} = (V_i, (E_i)_t \times RIW_t)$$

Mutation m is understood as a possibility of *new* edges e_i^* appearing in the next generation between players, that have not interacted before. As additional condition, the probability of a player starting such new connections is proportional to her PW, i.e. how well the players fare individually, so we have: $P(e_i^*) \sim PW(V_i)$. The better a player is suited, the more likely she will open up new interaction channels

$$m(G_1)_{t+1} = (V_1, (E_1)_t + \#e_i^*)$$

Selection s is understood as negative selection, i.e. interactions that disappear e_{sel} and do not survive to the next generation. An interaction disappears, if one of both players plays the

$\neg$ strategy, therefore, the number of selected edges is proportional to the number of actual instantiations of the $\neg$ strategy at a certain generation.

$$s(G_1)_{t+1} = (V_1, (E_1)_t - \#e_{sel})$$

In figure 10, these assumptions are summarized. We observe two graphs and their respective cohesion is given by their CI, which is at the same time an indicator for their internal ‘fitness’ (see footnote 5), since it counts the number of interactions within a graph. The two kinds of weightings that we developed in the last sections (RIW and PW) are implemented, in order to show how micro-level effects (how much utility a player gains, why a player stops interacting etc.) have a robust effect on the macro-level, i.e. the development of the population as a whole through time

For the sake of the argument, let us assume that, on the long run, selection will favor populations implementing a high inner cohesion. Therefore, the higher the cohesion index $CI(G_i)$, of a particular graph, the higher its internal ‘fitness’. This is relatively straightforward, since evolutionary fitness is given in terms of a variants’ relative frequency in a population (of organisms, features or traits), and $CI(G_i)$ is a measurement of the graph’s internal edges relative to the absolute number of possible edges. Recapitulating equation (2), it is given as:

$$CI(G) = \frac{CI_{abs}(G)}{max\ CI_{abs}(G)} = \frac{\frac{\#E_G}{\#V_G}}{\frac{\#V_G - 1}{2}}$$

Let us assume that a graph increases its’ fitness, if it has on average more internal than external interactions. So, for a graph to *increas*e its internal fitness in the next generation, it has to hold that:

$$CI(G)_{t+1} >^{!} CI(G)_t$$

If this condition is violated, a graph *loses* fitness. In combining our cohesion index and tools of population dynamics (figure 10), we can derive the following formula:

$$CI(G)_{t+1} = \frac{\#E(G_t)\times(e_i(G_t)\times t)+(-\#e_i(G_t))+\#e_i^{*}(G_t) \Big/ \#V(G_t)}{\frac{\#V(G_t)-1}{2}}$$

In words: The cohesion of the population of the next generation is given as an *index* (a relation) *of* the sum of those population's edges that are actually reproduced, those edges that newly appear and those edges that became selected *and* the maximal possible number of edges, given a fixed set of nodes. We arrived at this formula (9) by simply replacing $\#E_G$ in (8) with the terms for reproduction, selection and variation. For any distribution of edges and nodes, the dynamical analysis will yield a clear result.

Presenting this new model of how the cultural micro- and macrolevel can be integrated and to tell a story of how population structures could evolve through time, with finally *evolution selecting those groupings that yield a higher inner cohesion*, was the main contribution of this paper. Again, its goal was to communicate ideas, not a strict mathematical proof. The potential area of empirical applications is huge – reaching from linguistics, economics and other disciplines that do involve research on social networks or community discovery – and future work will probably inquire this space of possible implementations. The key question for empirical application is of course how the specific strategies will be interpreted and observed. Cultural evolution is a bundle of hugely complex systems evolving on different levels and much work lies still ahead.

__Table 5__: Ingredients of the population dynamical model for evolutionary game networks as structured graphs.

Competing Variants	$G_1 = (V_1, E_1)$	$G_2 = (V_2, E_2)$
Cohesion at t	$CI(G_1)_t$	$CI(G_2)_t$
Reproduction r(G$_i$)	$r(G_1)_{t+1} = (V_1, (E_1)_t \times RIW_t)$	$r(G_2)_{t+1} = (V_2, (E_2)_t \times RIW_t)$
Selection at t+1 s(G$_i$)	$s(G_1)_{t+1} = (V_1, (E_1)_t - \#e_{sel})$ with $\#e_{sel} \sim \#\neg$ strategies at t+1	$s(G_2)_{t+1} = (V_2, (E_2)_t - \#e_{sel})$ with $\#e_{sel} \sim \#\neg$ strategies at t+1
Mutation m(G$_i$)	$m(G_1)_{t+1} = (V_1, (E_1)_t + \#e_i^*)$ with $\#e_i^* \notin G_1$, and $P(e_i^*) \sim PW(V_i)$	$m(G_2)_{t+1} = (V_2, (E_2)_t + \#e_i^*)$ with $\#e_i^* \notin G_2$ and $P(e_i^*) \sim PW(V_i)$
"Reproductive-Interaction Weighting" (RIW) of e$_i$ at t+1	$RIW_t = e_i(G_i) \times t$	
Player's Weighting (PW) of V$_i$ at t	$u_c^{t_n}(V) = \left(u_c^{t_1}(V) + u_c^{t_2}(V) + \cdots + u_c^{t_n}(V)\right)$ with $u_c(V) = \sum_{1\leq n}^{n} \left(u_1(V), u_2(V), \dots, u_n(V)\right)$	

References

Allen, B. et al., 2017. Evolutionary dynamics on any population structure. *Nature 544*, pp. 227-230.

Aunger, R. e., 2000. *Darwinizing Culture: The Status of Memetics as a Science.* New York: Oxford Univ. Press.

Axelrod, R., 1984. *The Evolution of Cooperation.* New York: Basic Books.

Baraghith, K., 2020. Investigating Populations in Generalized Darwinism. *Biology & Philosophy 35/19*.

Boyd, R., Bogerhoff-Mulder, M., M., D. W. & Richerson, P., 1997. Are Cultural Phylogenies Possible?. In: Weingart, Mitchell, Richerson & Maasen, eds. *Human by Nature: Between Biology and the Social Sciences.* New York: Erlbaum, pp. 355-386.

Boyd, R. & Richerson, P. J., 1985. *Culture and the Evolutionary Process.* Chicago: The University of Chicago Press.

Dal Bó, P. & Fréchette, G., 2019. Strategy Choice in the Infinitely Repeated Prisoner's Dilemma. *American Economic Review 109 (11)*, p. 3929–3952.

Débarre, F., Hauert, C. & Doebeli, M., 2014. Social evolution in structured populations. *Nat. Commun. 5 3409*.

Dennett, D., 2018. *From bacteria to Bach and back.* New York: Norton & Company .

DiFrisco, J., 2018. Kinds of Biological Individuals: Sortals, Projectibility, and Selection. *British Journal for Philosophy of Science*, pp. 1-31.

Ereshefsky, M., 2017. Species. *The Stanford Encyclopedia of Philosophy*, p. https://plato.stanford.edu/archives/fall2017/entries/species.

Franke, M. & Wagner, E., 2014. Game Theory and the Evolution of Meaning. *Language and Linguistics Compass 8 / 9*, pp. 359-372.

Henrich, J., Boyd, R. & Richerson, P., 2008. Five Misunderstandings About Cultural Evolution. *Human Nature 19.2*, p. 119–137.

Hey, J., 2001. The Mind of the Species Problem. *Trends in Ecology and Evolution 16*, pp. 326-329.

Hofbauer, J. & Sigmund, K., 1998. *Evolutionary Games and Replicator Dynamics.* Cambridge : Cambridge University Press.

Huttegger, S., 2008. Evolutionary Explanations of Indicatives and Imperatives. *Erkenntnis*, pp. 409-436.

Jagers op Akkerhuis, G., 2016. *Evolution and Transitions in Complexity: The Science of Hierarchical Organization in Nature.* New York: Springer.

Killingback, T. & Doebeli, M., 2002. The Continuous Prisoner's Dilemma and the Evolution of Cooperation through Reciprocal Altruism withVariable Investment. *the american naturalist 160/4*, pp. 421-438.

Kleinberg, J., 2004. The Small-World Phenomenon and Decentralized Search. *SIAM News, Volume 37, Number 3*.

Lee, Y., Collier, T., Taylor, C. & Stabler, E., 2005. The Role of Population Structure in Language Evolution. *Proceedings of the 10th International Symposium on Artificial Life and Robotics*.

Lewens, T., 2015. *Cultural Evolution - Conceptual Challanges.* Oxford: Oxford University Press .

Lewis, D., 1969. *Convention - a philosophical study.* Cambridge, MA: Harvard University Press.

Lewontin, R., 2005. *The Wars over Evolution,* http://www.nybooks.com/articles/archives/2005/oct/20/the-wars-over-evolution/: New York Review of Books, 20 October .

Lieberman, E., Hauert, C. & Nowak, M., 2005. Evolutionary dynamics on graphs. *Nature 433*, pp. 312-316.

Mace, R. & Jordan, F., 2011. Macro-Evolutionary Studies of Cultural Diversity. *Philosophical Transactions of the Royal Society B 366*, pp. 402-411.

Maynard Smith, J., 1974. The Theory of Games and the Evolution of Animal Conflicts. *Journal of Theoretical Biology 47*, pp. 209-221.

Maynard Smith, J., 1982. *Evolution and the Theory of Games.* Cambridge: Cambridge University Press.

Mayr, E., 1963. *Animal Species and Evolution.* Cambridge: Harvard University Press.

Mesoudi, A., 2011. *Cultural Evolution – How Darwinian Theory can Explain Human Culture and Synthezise the Social Sciences.* London: The University of Chicago Press.

Mesoudi, A., forthc.. Blind and incremental or directed and disruptive? On the nature of novel varia-tion in human cultural evolution. *American Philosophical Quarterly 58/1.*

Millstein, R., 2010. The Concepts of Population and Metapopulation in Evolutionary Biology and Ecology. In: *Evolution Since Darwin: The First 150 Years.* Sunderland: Sinauer, pp. 61-68.

Mühlenbernd, R. & Franke, M., 2014. Meaning, Evolution, and the Structure of Society. In: *Proceedings of the European Conference on Social Intelligence (ECSI-2014).* s.l.:CEUR Workshop Proceedings 1283, p. 28–39.

Nowak, M., 2006. *Evolutionary Dynamics: exploring the equations of life.* Cambridge (Mass.); London : Harvard University Press.

Ridley, M., 2004. *Evolution.* Oxford: Blackwell Publishing.

Rossetti, G., Pappalardo, L., Pedreschi, D. & Giannotti, F., 2017. Tiles: an online algorithm for community discovery in dynamic social networks. *Machine Learning 106*, p. 1213–1241.

Scapoli, C. et al., 2005. Surnames and dialects in France: Population structure and cultural evolution. *Journal of Theoretical Biology, 237/1*, pp. 75-86.

Schurz, G., 2011. *Evolution in Natur und Kultur.* Heidelberg: Sprektrum akademischer Verlag.

Simon, H., 2002. Near Decomposability and the Speed of Evolution. *Industrial and Corporate Change, Volume 11, Number 3*, pp. 587-599.

Skyrms, B., 2010. *Signals: Evolution, Learning, and Information.* Oxford: Oxford University Press.

Steele, K., 2014. Choice Models. In: N. Cartwright & E. Montuschi, eds. *Philosophy of Social Science - A New Introduction.* Oxford: Oxford University Press, pp. 185-207.

Szathmáry, E. & Fernando, C., 2010. Chemical, Neuronal, and Linguistic Replicators. In: M. Pigliucci & G. Müller, eds. *Evolution - the extended synthesis.* Cambridge, MA.: MIT Press, pp. 209-249.

Tehrani, J. & Collard, M., 2013. Do Transmission Isolating Mechanisms (TRIMS) influence cultural evolution? Evidence from patterns of textile diversity within and between Iranian tribal groups. In: *Understanding Cultural Transmission in Anthropology: A Critical Synthesis.* Berghahn: s.n., pp. 148-164.

Tëmkin, I. & Eldredge, N., 2015. Networks and Hierarchies: Approaching Complexity in Evolutionary Theory. In: Serrelli & Gontier, eds. *Macroevolution. Interdisciplinary Evolution Research, vol 2.* Cham: Springer, pp. 183-226.

Wagner, E., 2012. Deterministic Chaos and the Evolution of Meaning. *British Journal for the Philosophy of Science 63 /3*, pp. 547-575.

Weibull, J., 1995. *Evolutionary Game Theory.* Cambridge/MA: MIT Press.

Zollman, K., Bergstrom, C. & Huttegger, S., 2013. Between cheap and costly signals. In: B280, ed. *The evolution of partially honest communication.* London: Proceedings of the Royal Society.

Chapter 10

Metaphysics of Evolution: Ontology and Justification of Generalized Evolution Theory

Gerhard Schurz[i]

Abstract

Section 1 introduces generalized evolution (GE) theory. Here, the three core principles of the theory of evolution − reproduction, variation and selection − are detached from their biological basis, abstracted and extended to other domains, in particular to the domain of cultural evolution (CE). Section 2 investigates the ontological foundations of GE and CE theory. They consist in entities and structures that must be realized to get the three modules of evolution running. These entities include self-reproducing systems with variation and populations of such systems in environments with limited resources. A major difference between the reproductive units of BE and CE consists in the fact that genes act as constructors of evolutionary systems, while memes play the role of modificators. Section 3 is devoted to the *abductive justification* of Ge and CE theory that proceeds in three step: First, CE theory is justified, second, based on the first step GE theory is abductively justified as the common core of Be and GE, and thirdly, the fruitfulness of GE theory is demonstrated by its applicability to further domains.

Keywords : generalized evolution theory, cultural evolution theory, ontology of evolution, genes, memes, evolutionary systems, abductive justification.

---

[i] Heinrich Heine University Düsseldorf, Germany, Department of Philosophy.

10.1. Introduction

Until the 1970s the groundbreaking successes of the theory of evolution had been largely restricted to work in the biological sciences. In the meantime, a substantially more general research program has been established, known as *generalized evolution* (GE) theory or generalized Darwinism. The main philosophical idea that unites the many-faced approaches found in the general research program is that human culture obeys similar developmental principles to those characterizing biological systems. In contrast to sociobiology (Wilson 1975), GE theory doesn't reduce evolution to the biological-genetic level. Instead, a distinct level of *cultural evolution* is postulated based on the transmission of information acquired through social learning. Within GE theory, the three Darwinian core principles of the theory of evolution are detached from their biological basis, abstracted and extended to the domain of *culture in the broad sense*, including language, law, science and technology. The three principles of GE theory can be explicated as follows:

> *Reproduction*: Certain entities, the 'evolutionary systems,' reproduce themselves and pass on certain 'inherited' traits. Every reproduction cycle produces a new generation.
>
> *Variation:* Reproduction brings along new variants that are inherited at the same time.
>
> *Selection:* Certain variants are fitter under the given environment, that is, they have the disposition to reproduce faster than others, replacing other variants in the long run.

Several authors have formulated these three principles in similar ways (e.g., Sober 1993, 9; Dennett 1995, 64 f.; Hodgson and Knudsen 2006). Of course, formulations are different. For example, some authors prefer the notion of "retention" over "selection" (Campbell 1974; Reydon and Scholz 2015). The difference between mere retention and reproduction is clarified in sec. 2.1 below, where we will argue that retention without reproduction enables only a form of proto-evolution. Other authors introduce the condition of "inheritance" as primitive; but note that "reproduction" entails "inheritance" and condition 2.) requires that variations are inherited. Moreover, the concepts of selection and fitness can be understood in different ways that will be reflected in section 2.5.

Already at the end of the 19th century there had been attempts to generalize the theory of evolution from biology to the social sciences and humanities, for example in Spencer's

(1851) 'Social Darwinism.' These generalizations were largely based on non-Darwinian and pre-scientific understandings of the notion of evolution. In particular, there is in Darwinian evolutionary theory neither an automatism for a 'development towards higher forms' nor a 'law of the selection of the strongest', as assumed by Spencer (1851). The major scientific impulses for the development of a GE theory began in the 1970s. Cavalli-Sforza and Feldman (1973, 1981) and later Boyd and Richerson (1985) argued that cultural evolution has to be treated as a second dimension of evolution, thereby framing evolution as a *dual inheritance system* (cf. Mesoudi et al. 2006). At the same time, evolutionary game theory was developed by pioneers such as John Maynard Smith (1974, 1982). Prominently, Richard Dawkins (1976) developed the concept of the *meme* as the cultural counterpart of the gene. Memes are human ideas and abilities acquired by learning and reproduced by mechanisms of cultural tradition.[1]

According to these accounts, evolution in humans takes place on two levels: biological evolution (BE) at the level of genes and cultural evolution (CE) at the level of memes. The generalization of the evolutionary principles to the level of CE is not meant as a mere 'analogy' or 'metaphor', but is taken literally, because the three Darwinian principles apply at both levels.

First, on both levels evolution is based on a corresponding kind of *evolutionary systems*. In BE these systems are biological organisms in interaction with their natural or social environment. Remarkably, also in CE the primary evolutionary systems are biological organisms, namely human individuals and their social, technical and cultural systems.[2] This points rests on the fact that CE is piggybacked by BE (see sec. 2.2 below). BE studies humans primarily by genetically inherited phenetic properties, while CE studies humans primarily by culturally (non-genetically) acquired properties.

Moreover, on both levels there is *reproduction*: the evolutionary systems at both levels possess certain subsystems which are primarily replicated or reproduced from each other; these

---

[1] Two further developments were (i) the generalized evolution theory of Donald Campbell (1960, 1974), who was also one of the originators of evolutionary epistemology (together with Lorenz 1941/42 and Popper 1979) and (ii) the idea of Generalized Darwinism in economics (Nelson and Winter 1982, Hodgson and Knudsen 2010).

[2] We ignore here the existence non-human animals being capable of limited cultural evolution.

subsystems are called *reprones* (Schurz 2011, sec. 6.1). The reprones of BE are genes and genotypes (sometimes combined with epigenetic activation patterns; cf. Jablonka and Lamb 2005, ch. 4). The reprones of CE are memes, that is, acquired information structures which are stored in human brains.

There are also several important differences between BE and CE (Schurz 2021, sec. 3). Gene reproduction is based on replication, that is, on a physical or 'syntactic' copying process. In contrast, memetic inheritance is based on semantic reproduction, in which a phenetic expression, typically a behavior or a spoken or written text, is reconstructed by the meme-learner (Sperber 2000, 165-167; Schurz 2011, sec. 9.5). BE is characterized by the additional complication of sexual reproduction and genetic diploidy, a peculiarity that is not found in CE. Instead, what frequently occurs in CE is blending inheritance (Boyd and Richerson 1985), i.e., the acquisition of a combination of memes learned from different 'cultural parents.'

Variation arises in BE in the form of mutation and recombination of genes, and in CE through the interpretation and modification of acquired memes. *Selection* occurs simply because in the given environment, evolutionary systems with certain reprones reproduce faster than systems with different reprones. This leads to selection, because the occupation number is limited in BE as well as in CE: Only a finite number of organisms may coexist in a non-destructive way within a given habitat, and only a finite number of memes may coexist in a coherent way within one human brain.

Those traits of an evolutionary system that are produced from the reprone during its normal development are called the phenetic traits or *phenes* of the evolutionary system. The phenes of BE are the parts and capacities of the biological organism, including capacities for niche construction (Odling-Smee 2010). The phenes of CE are the technical or cultural products, acquired skills or dispositions, and social institutions.

There is a further and more subtle difference between the reprones of BE and of CE, that has to do with the fact that CE piggybacks on BE. The genes in the diploid chromosomes of the germ cells act as *constructors* of the biological organism, of course in interaction with the environment, but nevertheless the whole biological organism develops 'from scratch'. In contrast, the memes of CE are introduced into already existing evolutionary systems; they act as *modificators* and *developers* of the informational control centers of these systems (human brains), but the systems in which the memes unfold themselves have been created within BE.

In this respect, memes are similar to 'viruses' in a generalized (not mainly malevolent but benevolent) sense. The metaphor of memes as 'viruses of the mind' has been suggested by Dawkins (1976, ch. 11; 1986), Dennett (1995, 47) and Blackmore (1999). The analogy has been heavily criticized and we cannot enter this debate (Schurz 2011, sec. 9.3). For sake of fairness one should be clear that there is something true about this analogy –namely that memes are not constructors, but modificators of already existing evolutionary system that 'intrude' by mechanisms of learning.

Summarizing, the theory of GE categorizes the basic components of generalized evolution and their respective correlates at the levels of BE and CE as follows:

***Table 1**: Correspondences between biological and cultural evolution.*

GE (Generalized Evol.)	BE (Biological Evolution)	CE (Cultural Evolution)
Evolutionary systems (e.s.)	Organisms in their environment	Humans and human societies (in their environment)
Reprones	Genes in the cell-nucleus, epigenetic patters	Memes (acquired information, software of the brain)
	Constructors of e.s.	Modificators of e.s.
Phenetic traits (phenes)	Organs, abilities, niche construction	Behavioral & cognitive skills, technologies, language, ideas & knowledge
Reproduction	Replication	Imitation & learning between humans
	DNA copy	Semantic reproduction
Inheritance system	sexual (diploid)	asexual (blending inheritance)
Variation	Mutation and recombination	Interpretation and variation of transmitted memes
Selection	Higher rates of reproduction of genes (& epigenetic)s	(due to different causes) of memes

10.2. The ontological architecture of generalized evolution theory

Reydon and Scholz (2015) complain that it is unclear what the ontological assumptions of GE theory really are. I propose that the ontological architectonics of GE theory follows the three evolutionary modules defining *evolutionary systems*. Due to its generality, the ontology architectonics of GE theory has to be sparse. We propose the following two *minimal* ontological assumptions of GE theory that care needed to get the evolutionary dynamics running:

1. *Self-reproducing systems with variations:* There must be self-reproducing systems. These are open systems that possess the capability of reproduction, which requires a constant energy and matter influx from their environment. The self-reproduction must not be perfect, but must produce variations that differ in their reproduction rate and are themselves reproduced.

2. *Populations* of self-reproducing systems *competing* for resources: Variants of evolutionary systems coexist in the form of *populations*, inhabiting an environment with upper limits for population numbers, which means that the variants are in some sort of competition. This implies that the 'fitter' variants – those who have the disposition to reproduce faster than others – tend to replace the less fitter variants, as long as the reproductively relevant environmental parameters remain stable.

10.2.1. Self-reproducing systems with variations

The most difficult condition for the emergence of evolutionary systems, is the formation of systems that have the capability of self-reproduction. Of, course, the self-reproduction must not be perfect but must involve inheritable variations, but this condition is easy realized, since there is always noise and disturbance in the word; the difficulty and highest barrier for the emergence of evolving structures the universe is self-reproduction. The reproduction of a system is higher demand than the reproduction of a mere pattern within a system, as for example in the growth of a crystal (see below). Systems are material ensembles delimiting themselves from their environment by spatial borderlines. Self-reproduction means that a new and sufficiently similar 'copy' of the old system is created or produced, at a nearby location of the old system. To avoid overpopulation the old system must die after a variable amount of time. A self-reproducing process requires a high amount of complexity of and energy-influx into the system.

General ontological descriptions of different kinds of systems are given in *generalized systems theory*, developed e. g. by Bertalanffy (1979) and Rapaport (1986).[3] This is the most general discipline in which GT theory is embedded. The description of the real world as a set of *systems* is based on two fundamental ontological properties of it: First, does not consist of isolated individuals, but the individuals entertain numerous *causal relations*. Second, *not* everything is (significantly) related to everything; the interconnected individuals of the world group to certain clusters resp. systems, that consist of strongly connected *parts,* but are to a certain degree *causally shielded* from their environment. Because of these facts, systems possess spatial *borders* towards their *environment* that are more-or-less stable over time and that grant the system a certain (non-perfect) *identity over time*, by which they are demarcated from their *environment.*

A fundamental system-theoretical distinction is that between *closed*, isolated and *open* systems (Schurz 2005). Closed systems have neither an exchange of matter nor of warmth with the environment; isolated systems just have an exchange of warmth. By contrast, in open systems a flow of both matter and energy occurs between system and environment. For closed systems, their self-identity over time is a matter of the absence of external influences due to contingent facts; if our planetary system were to be destroyed once by a giant meteor, then it remains destroyed and doesn't regenerate. In contrast, the stability of open system is based on their capability of *self-regulation*, which enables them to compensate interfering influences of the environment through countermeasures (as studied in *cybernetics*; tracing back to Ashby 1957). Their identity in time is preserved within a certain *normal range* of the system's states; if the systems remains outside this of normal range for too long, it 'dies' and lose their identity (Schurz 2001; Schurz 2011, sec. 7.2).

Evolutionary systems are, of course, open systems, as self-reproduction requires influx and outflux. Moreover, evolutionary systems are highly self-regulatory (including reproduction and self-organisation as a higher forms of self-regulation). In fact, most self-regulatory systems inhabiting our earth are either the result of natural evolution, or are designed by humans, being

---

[3] More on general system theory cf. Müller (1996) and Mesarovic and Takahara (1989).

the result of cultural evolution. This is not surprising, because self-regulation requires a high amount of complexity, whose emergence is made probable by continuous optimization through evolutionary selection. To give some examples: A stone lying on the earth (not impacted by anything) is a closed system; two stones lying side by side are (one 'set' but) two systems; a planet is an isolated system (as long as no large comet crashes into it); a tree is an evolutionary system and a leaf of the tree is a part of it; the leaf falling to ground turns into a non-evolutionary but still open system slowing converting into soil.

There are also some self-regulatory systems at the pre-evolutionary, physical or chemical level, but they are infrequent and less stable, because the forces of evolutionary 'maintenance selection' are missing. For example, the cycle of water that spends us regular rainfall is a self-regulatory mechanism, but it is highly vulnerable by dryness and desertification. Also the development of macroscopically ordered systems is often based on temporally limited self-regulatory process in which certain micro-patterns are preferred over others. In this sense, Campbell (1974, 55) described the process of crystal growth as an instance of his *variation-selection-retention model* of evolution: a certain geometric neighborhood configuration of ionic molecules is energetically preferred over other configurations; due thermal noise (Boltzmann processes) every molecule has enough time to find its optimal position, so that finally a macroscopic crystal symmetric emerges. The crucial difference of a variation-selection-retention process to proper evolution is that that the static retention of preferred patterns comes to a *standstill* after all parts of the system have found their optimal configuration. Schurz (2011, sec. 6.4-6.5) suggests calling this type of development *proto-evolution*. Also the formation of planetary systems is driven by a proto-evolutionary process (Ward and Brownlee 2000, 43–46).

In conclusion, *self-reproducing systems* are the most important subcase of self-regulatory open systems that have the ability to reproduce themselves, with heritable variations, and are the product of a long evolutionary history. The three most important realization levels of evolutionary systems are (1) the *biochemical* level of self-reproducing RNA macro-molecules, (2) the *proper biological* level of organisms, and (3 the *cultural* level of acquired information and abilities of biological systems that possess the capacity of *cultural* evolution; until today almost exclusively *humans*.

We finally turn to the ontological basis of the condition of *variation*. As explained, due to disturbances whose influence cannot be completely eliminated, it is natural that errors occur in the self-reproduction process. The non-trivial point is not the explanation of the existence of reproduction errors, but the maintenance of the error-rate within a suitable size, not too big and not too small, in order to get evolution running. The rate of variation of cumulative evolution must be neither too high, nor too low (cf. Dawkins 1999). If it amounts to e. g. one mutation per 1 million reproductions, evolution almost comes to a halt and the adaptation to altered environmental conditions occurs too slowly. If it amounts to one mutation per five reproductions, then every advantageous variant will quickly *randomize*, i. e. statistically disappear by successive random errors, before it can develop its fitness advantage, and no selection occurs. The mutation rate's optimum is dependent on the environment and lies somewhere in between (e. g. with mice in BE at about 10^5 mutations per gene). In CE, variations occur much more frequently than in BE, since every new generation varies tradition and creates new things. In some cultural domains variation is in fact too high to enable sustainable evolutions; many cultural variations randomize before they can develop a sustainable fitness advantage. Examples may be given in the cultural evolution of fashion (cf. Lumsden and Wilson 1981, 170f.), art, or speculative philosophy (cf. Schurz 2011, sec. 9.6.4).

10.2.2. Ontological specifics of cultural evolution

As explained above, CE is piggybacked by BE. The evolutionary systems of CE – humans and their cultural systems – are not constructed by the reprones of CE, the memes. Rather, they are taken over from BE and are merely modified and developed by CE. For this reason memes have been compared to 'viruses' (an analogy whose misleading potential is greater than its payoff).

This peculiarity of CE has itself evolutionary reasons. The most important *intermediator* between BE and CE is *individual proto-evolution*, by which I mean the development of the human brain by *individual* learning. Individual learning has been invented within BE, since not all advantageous adaptions can be encoded by rigid gene structure, in particular not adaptions to environmental changes within the individual's lifetime. A brain that has the ability to learn within the individuals lifetime, by *inductive conditioning* processes, is a clear evolutionary advantage. Trial-and-error learning can be considered as a kind of evolutionary process. Yet we call it proto-evolution, because it is not sustainable evolution, but comes to an end by the individual's death (Schurz 2011, sec. 11.1). Individual learning is made *sustainable* by the mechanism of social learning within CE, by passing learned information and behavior from one to the next generation. This is the way CE was evolutionarily set into play in the first place. And this explains why CE is piggybacked by BE and why the reprones of CE are not constructors but modificators of the evolutionary systems within which they act.

There has been an intense philosophical debate about the ontological nature of memes – whether they should be characterizes as 'software' structures (i.e.. non-genetically determined structures) of the human brain, as cognitive-behavioral dispositions, or as externalized information structures.[4] Certainly, all three possibilities of meme representation are important. In particular, for complex memetic systems the possibility of external storage in information media (books, electronic media) becomes crucial. Nevertheless we think that the understanding of memes as software structures of the brain has two advantages:

First, for a document or artifact to function as a meme, it needs subjects who reconstruct and understand the meme's meaning and use. An externalized meme can only become efficient and evolve when it enters a human brain, or more generalized, the information control center of some other evolutionary system. This point is decisive for measuring cultural reproduction rates, in distinction to mere growth rates; what counts are not numbers of external meme-copies, but numbers of human heads or brains having internalized these memes (cf. Ramsey and De Block 2017, 10).

Second, the meme-as-brain-software account allows for a clear distinction between memes qua reprones and cultural traits or phenes. Cultural phenes may be behavior, technological artifacts, oral utterances or writings; they are caused by brain-memes in interaction with the environment. Thus this distinction resolves a concern, raised by Hull (1982), that in the field of CE there would be no clear differentiation of memes and phenes.

---

[4] The unclear ontological status of memes has been criticized both by biologists (Jablonka and Lamb 2005), anthropologists (Sperber 2000) and philosophers (Reydon and Scholz 2015). Some CE theorists (Dennett 1995, 483; Durham 1991, Blackmore 1999) have characterized memes as acquired information that may also be stored outside the human brain. Other have proposed identifying memes with neuronal structures in the brain (Dawkins 1982, Aunger 2002; Schurz 2011, sec. 9.1). For further controversies cf. Aunger (2000), Gardner (2000), Cladière et al. (2014), Ramsey and De Block (2017).

Let us finally ask to which degree CE may become independent from BE. All what CE requires as necessary condition is that that there are sustainable collectives of complex systems that have (i) the ability to learn, to store the learned information and skills somehow, (ii) to vary these information structures by cognitive processes, and (iii) to reproduce this information by passing it to other systems of their population. Even if ordinary biological reproduction would stop and be displaced by the perfect cloning of one or of some privileged humans, the so-created population of clones would be able to entertain cultural evolution. Moreover, even if these biological closes were replaced by robots, CE could continue. On the other hand, the need of sustainable duration makes it plausible that the systems carrying CE (be they biological systems or robots) have to be themselves functionally maintained by some sort of evolutionary maintenance process, i.e., not only the memes within the evolutionary systems but also these systems themselves should be subject to evolution. Finally, note that I reject these biotechnological possibilities because of the danger of their misuse; playing through these thought experiments makes me feel like a mini-Oppenheimer telling the public about the possibility of an atomic bomb.

10.2.3. Ontological specifics of biochemical evolution

It might be thought that biochemical evolution is essentially different from proper biological evolution, because biological organisms are 'living' beings, while macromolecules aren't 'living'. On a closer look this rigid difference disappears. At the end of the 1960s it was understood how from major chemical components of the primal Earth’s atmosphere, through lightning or spark discharges, organic compounds can arise from which spontaneously shorter RNA chains could form (Maynard Smith and Szathmáry 1996, 31, 73). However, Eigen (1971) showed that because of the high rate of copying errors, RNA chains can only be replicated up to a length of about 100 bases; beyond this point randomization begins. For longer RNA chains to be able to replicate, suitable *enzymes* are needed. In current cells the enzymatic function is taken over by proteins, but proteins are themselves produced by RNA; so how could long RNA-chains be formed at the beginning of evolution? ('Eigen's paradox'). The decisive step for the solution of this riddle was the discovery that certain RNA chains have themselves an enzymatic function, thereby helping other RNA-chains to replicate. In conclusion, the beginning stage of the prebiotic evolution was a pure RNA world, and RNA-chains had to form suitable *cooperation* structures, in which some RNA-chains are helping other RNA-chains to replicate themselves. The well-known problem of evolutionary cooperation is that cooperative structures may be undermined by *egoists* (Sober and Wilson 1998). RNA-chains that enjoy being helped but don't help others, and by acting egoistically, replicate faster than altruistic RNA-chains. The first proposal for a solution was Eigen's *hypercycle* based on reciprocal cooperation (Eigen und Schuster 1977). This diminishes the frequency of egoists, but doesn't exclude egoistic intruders that may destroy the hypercycle. A second and more stable solution consists in encasing a hypercycle in a *membrane*, in order to prevent the intrusion of parasites (Maynard Smith and Szathmáry 1996, 53 ff.). This may happen spontaneously based on membranes of lipid molecules adhering to underwater volcanic rocks and spontaneously forming closed vesicles; if a hypercycle is located below the lipid layer, it is encased in the vesicle. In this way one conjectures that the first protocells have originated; of course, the ability of cell division requires additional mechanisms requiring suitable interactions between the RNA chains in the nucleus and the cell membrane (ibid., 103). Note that the emergence of the first primitive cells was the first historical time in which the specialization of evolutionary systems into reprones and a surrounding system took place.

10.2.4. Populations of self-reproducing systems competing for resources

'Populations' are a central concept of biological Darwinism. In biology a population is traditionally understood as a group of interbreeding species members living together in a habitat (Godfrey-Smith 2009); but 'interbreeding' can hardly be conceived as a necessary condition for CE, even not in a suitably generalized understanding of it. In biology, a notion of populations weaker than that of a group of interbreeding individuals has been developed by Millstein (2010). It is based on the idea that the members of a population stand in regular causal interactions of either a reproductive or competitive or cooperative kind (for similar ideas cf. Godfrey-Smith 2009, 39; Hodgson and Knudsen 2010, 34). Although this broader concept of population has applications in CE, it is questionable whether all kinds of CE involve populations of this kind. Some authors have doubted whether populations are really a necessary ontological assumption of GE (Bouchard 2008). Biological evolution requires *many* individuals because within each individual only *one* genetic variant can exist; the competition of genes requires many individuals. In contrast, in cultural evolution many memes can compete within one individual, and the individual may select those that turn out as fitter in interaction with the environment. We can even devise a thought experiment in which one individual clones itself, passes its favored memes to its clone, dies thereafter, and in this way implements CE within one lineage, consisting of just one individual per time. Although such a scenario is possible, it is highly improbable. A system that produces just one clone of itself is exposed to a high danger of dying out by disturbances. Only lineages producing many clones at one time will have realistic survival chances. As soon as this is the case, there are populations in the minimal sense of systems coexisting in the same environment and competing for its resources, thereby developing different meme-structures that lead to different success rates in the reproduction these memes.

In conclusion, what seems to be required for making evolution running is the existence of a group of reproducing evolutionary systems inhabiting the same environment and being in mutual competition. Specific kinds of evolution may require additional sorts of interactions. For example, proper biological evolution requires sexual reproduction; proper cultural evolution requires inter-communication; group evolution requires cooperation. However, these interactions are not mandatory conditions of GE theory. What is mandatory is the effect of this competition: Certain variants have the disposition to reproduce faster than others and since the limited resources of the environment sets upper bounds to the population density, the fitter variants replace the less fitter variants in the long run.

10.2.5. The functional notions of fitness and selection

It is conspicuous that we never attempted to characterize 'fitness' in a content-related, substantial way. We rather characterized the fitness of a variant, in a competing ensemble of variants, by is reproduction rate in the given environment. More precisely, fitness is not identified with the *actual* reproduction rate of a variant, but with its probabilistic *disposition* (also called the *propensity*) to reproduce in a certain rate, because survival by mere luck doesn't count as fitness. In conclusion, we regard it as a *conceptual truth* of GE theory that the *fitness of a reprone* R in a given type of environment E is proportional to the numerical propensity of R to reproduce, defined as the expected number of R's offspring in in environment (cf. Schurz 2001; 2021, sec. 3.4).

Many evolution theorists define fitness in this way (Futuyama 1998, Ridley 1993, Sober 1993, De Jong 1994, 6). Philosophers have objected against this way of explicating fitness, but we think their objections are unjustified. For example, Popper (1979, 83 f) and his followers have complained against BE theory that it identifies fitter variants with those having a higher rate of reproduction and thereby turns the 'central law' of evolution, the *survival of the fitter variants*, into a tautology. Also in the domain of CE, several authors (e.g., Wilson 1999, p. 206; Henrich et al 2008) have objected that the notion of cultural fitness is unclear, for the reason that cultural fitness cannot be defined independently from cultural reproduction rates, which would make the 'law of evolution' tautologous. Several authors attempted to avoid the tautology accusation and tried to characterize fitness independently from reproduction rate, as certain *causes* of high reproduction rates (e. g. Vollmer 1988, 260). We think that these attempts rely on a misunderstanding of evolution theory. A definition of fitness that is independent from reproduction and at the same time suitably general cannot exist, even not in the biological domain and even less in the cultural domain. The fitness of given traits is *relative* to the selection conditions of the respective environment: what is fitter in one environment is less fit in another environment (cf. Weber 1998, 202). For example, neither "strength", nor "complexity" nor "intelligence" are general biological fitness factors; with a little background one can easily imagine scenarios in which the less strong, complex or intelligent organisms are reproductively more successful in the long run. Therefore, a general definition of fitness can only be given through its identification with the disposition to reproduce in a certain rate.

In defining fitness as a disposition we assume the standard view of dispositions, according to which a disposition of an object x consists in a nomological regularity of x to produce a certain reaction, provided a certain initial (or trigger) condition is realized (cf. Prior et al. 1982) – either with certainty, or with a certain probability. In this view, a disposition is not a structural or categorical, but a *functional* property, that can be materially realized in different ways. For example, elasticity can be realized by different 'causal bases', i.e. molecular structures – compare the elasticity of a rubber band with that of metallic spring. There is an alternative view (e.g., Quine 1960, Mumford 1998) that identifies a disposition with its causal base. I think this view is refuted by the fact that, typically, different dispositions have the same molecular structure as their causal basis (Schurz 2014, 138-140). This question is highly relevant to the debate on fitness: For example, Mills and Beatty (1979, 271) define fitness as the disposition to reproduce, but argue for a non-standard view of dispositions that identifies fitness with causal properties of the organisms. For the explained reasons such a characterization cannot work in a sufficiently general way. In contrast, we understand the disposition to reproduce as a functional property, that may be realized by different causal bases.

In conclusion, that fitness has no reproduction-independent definition is a general feature of GE theory and doesn't constitutes an argument against it. This insight implies that the 'law the survival of the fittest' is indeed a mathematical-analytic truth, that roughly speaking has the following form: "*If variants, variation rates and fitness values are such and thus*, then iterative evolution will lead to this-and-this result" (cf. Sober 1993, 71). However, this fact does not imply that the theory of evolution becomes empirically vacuous. Instead, its factual content is contained in the *if-condition* of this analytic conditional: that certain observable entities *are* evolutionary systems whose variants, variation rates and fitness values are such and such. These assertions are enormously strong in empirical content (cf. Schurz 2011, sec. 7.5).

Precisely the same considerations apply to the scientifically advanced notion of *selection*. Selection is a *byproduct* of variations in fitness, i.e., reproduction rates, in an environment that sets upper limits to the number of variants that may inhabit it. Thus, like fitness, also selection is a *functional* property, a disposition of a population in a given type of environment. We emphasize this because recent criticisms of the notion of selection in CE theory, in particular that of Lewens (2015), rest on the misunderstanding that selection means a particular causal process, meaning selection in the sense of selecting variants that are more efficient in regard to one parameter such as speed, strength or intelligence. In contrast, selection in the functional sense may have variable causal bases; diverse mechanisms in the dissemination and medial presentation of memes may be decisive for the cultural attractivity of a meme.

10.3. Abductive Justification of CE and GE Theory

How can GET in general and CET in particular be justified or confirmed, from a standard philosophy of science perspective? Scientific theories involve, besides observable concepts, theoretical concepts that are not directly observable (Schurz 2014, ch. 5). In GE theory, for example, 'reproduction rate' is an empirical concept while 'reprone' is a theoretical concept. Theoretical concepts cannot be introduced by a simple inductive generalizations. Rather, scientific theories are typically justified by methods of *abduction*, or *inference to the best explanation* (IBE). Abductive inferences go back to Peirce (1878); Harman (1965) transformed

Peirce’s concept of abduction into the schema of the inference to the best explanation. It can plausibly be argued that abduction in the understanding of Peirce is synonymous with Harman’s notion of an IBE (cf. Douven 2011, Niiniluoto 2018, sec. 1.4). Therefore, in the following we will use the notions of ‘abduction’ and ‘IBE’ synonymously. They can be represented as follows (cf. Lipton 2004, Schurz 2008):

General schema of an abductive inference:

Premise 1: A (singular or general) fact E that is in need of explanation.

Premise 2: An epistemic background system S, which implies for some hypothesis H that H is the best explanation for E available in S.

Conclusion (explanatory conjecture): H is true.

Using the abductive strategy, CE theory and GE theory can be justified in three steps.

First: CE theory has to be justified as the best available explanation of the phenomena of cultural evolution. A sketch of this is developed in this section.

Second: If step 1 is successful, GE theory can be justified as the *unifying core* theory that is common to biological as and cultural evolution. The major unification achieved by GE theory lies in

(a) providing the common ontological foundations of evolutionary systems, as outlined in the previous section, and

(b) describing the general dynamical laws of evolution within general population dynamics (GPD). As demonstrated in Schurz (2021), GPD's equations subsume the laws of biological population dynamics (Fisher 1930, Ridley 1993) as well as those of evolutionary game theory (Weibull 1995, Skyrms 2014). Hand in hand with GPD's dynamical laws goes a systematic classification of different parameter settings, such as frequency-independent versus frequency-dependent selection, reflexive versus interactive frequency-dependence, and positive versus negative reflexive frequency-dependence (Schurz 2021, sec. 4.2).

Third: The fruitfulness and unification power of GE theory is corroborated by finding further applications of it. One such example, evolution processes at the biochemical level, have been described in section 2.4. Further applications have been found, for example, in *immunology* (Cziko 2001, 19) and in *neural growth* (Edelman 1987).

The major justification load for GE theory, and most controversial point, is the justification of CE theory. To justify CET abductively, as an inference to the best *available* explanation, we have to ask: what are the major competitors to CET in regard to the explanation of human cultural development? For this purpose, we have to compare CET with its major competitor and demonstrate the advantages of the explanations based on CET.

The major competitors of CET are often their major critics. Some biologists have objected that CE is not sufficiently similar to BE and not more than a loose metaphor (Gould, 1991, 65; Maynard Smith, 1986). Others have argued that the evolutionary aspects of human development can be better explained by sociobiology (Wilson 1975). So the first major competitor to CE theory is *sociobiology*.

Likewise, sociologists and cultural anthropologists have objected that CE is not sufficiently similar to BE, but they infer from this fact the opposite conclusion that cultural evolution should be explained by radically *non-evolutionary accounts* and regard CE theory as an "imperialistic" attempt at reducing cultural studies to natural science (Bryant 2004, Perry & Mace 2010). There are two main alternatives to CET arising from the humanities and social sciences:

Individual-centered accounts are exemplified in action theorists and rational choice theorists, which intend to explain history and cultural development in terms of the intentions of rational agents (Dray 1957; Voss & Abraham, 2000, Steele 2014).

Society-centered accounts (Black 1995) attempt to explain cultural development by the autonomous laws governing history or society, that are neither reducible to the biology nor the psychology of the individuals.

This classification of alternative accounts to CE theory is course-grained and patchy. We cannot give a comprehensive treatment of alternatives, but understand the following arguments as a pars-pro-toto discussion, illustrating how abductive justifications work.

10.3.1. Arguments against sociobiology

Sociobiology intends to explain CE by reducing the essential patterns of CE to BE. The possibility of such a reduction is highly implausible on several counts. First, because of the substantially higher speed of CE in comparison to BE (Boyd & Richerson 2000). Humans and chimpanzees are as genetically similar as rats and mice or lions and tigers; how then could humans progress culturally as far away from the chimpanzees in only 2 million years, as has never occurred elsewhere in evolution between genetically so strongly related genera? Only the assumption of an independent level of cumulative CE, whose speed exceeds that of BE by several decimal powers (10^3 to 10^4) can explain this mystery. Moreover, sociobiology explains cultural variation by different environmental conditions that provide different inputs to the same genetic endowment. If such explanations were possible, then the same environmental conditions would have to lead to the same or at least similar cultural developments (since all humans are comparably closely related genetically). However, a number of cross-cultural studies provide examples of social groups, which in spite of an equal genetic endowment and equal natural environmental conditions have developed cultural life forms that are sustainably different (cf. Salomon 1992, Inglehart 2003, Schurz 2007). Sociobiologists have recognized cultural variations of this sort (Lumsden and Wilson 1981), but they view it as kind of phenotypic plasticity or 'noise' and cannot explain their evolutionary robustness and sustainability.

10.3.2. The argument against action and rational choice theory

According to rational choice theory, people have an innate tendency to do what is best for them. Some authors drawn from that the conclusion that CET can finally be reduced to rational choice theory (e.g., Schatzki 2002). Can CE be reduced to the achievement of individuals, as intended by ‘methodological individualism’ of individualistic action or rational choice theory? The answer seems to be *no*, since CE transcends million fold what individual geniuses can achieve. Nothing can illustrate this better than the evolution of science and technology. The inventor of the first car wouldn’t have dreamed of how a present-day Mercedes looks like, and it is similarly true for the inventor of the first refrigerator, television set, etc. Imagine as a thought experiment that from one generation to the next all so far accumulated information and technical appliances would abruptly be destroyed. The following generation would be transferred back to the level of knowledge and technology of people in the Stone Age (many thousand years ago), and it would require thousands of years to make up for this gap again. Conversely, when catapulting a toddler from the few tribal societies on earth (e.g., in New

Guinea) that currently still live on the Stone Age level into the Western civilization, the child will effortlessly make up for this time and jump of hundreds of thousands of years within 20 years of education.

We do not deny that action theory can offer satisfactory explanation of *short* historical episodes, but it cannot explain *long-term* developments in CE that have been intended by nobody. Basalla (1988) has demonstrated the non-intentionality of CE using many examples from technological evolution, and similar points are made by Diamond's (1998) evolutionary history of humankind. The long-run result of cumulative CE is typically not explainable in terms of human intentions or plans, but can be explained as the cumulative effect of iterated selection in CE. Gorbachev wanted the opening of the Communist bloc towards the West and got it, but his intention was certainly not the collapse of the Soviet-Communist bloc, just as little as Marx had wanted Stalinism or Einstein had wanted a-bombs.

10.3.3. The argument against pure sociology

Are there irreducible macro-laws governing the structure and development of societies? The view that there are historical laws that act like laws of nature and determine history has been defended, for example, within Marxism. This view is called *historicism* and has been thoroughly criticized by many authors, in particular by Popper (1957). Proponents of a “pure sociology” (going back to sociologists such as Simmel or Durkheim) have postulated irreducible sociological laws. In the centre of their theories are sociological learning or role accounts, that regard the structure of the society as only major cause of cultural development. However, purely sociological explanations are not really explanatorily powerful: they cannot explain why in CE exactly *these* and no other cultural structures have obtained. For example, pure sociological accounts can hardly explain why in more-or-less all societies there is a certain amount of normative regulation, of cooperation and, of division of labour. Since purely sociological accounts explain the action of humans by the structure of society which in turn is the effect of these actions, they are caught in a complete explanatory circle. Note that our objection is aimed only at radical versions of purely sociological explanations. In contemporary sociology there are also 'pure' sociologists who attempt to explain variation in culture and conduct ethnographic micro-analyses.

Funding

This work was supported by the DFG (Deutsche Forschungsgemeinschaft), research unit FOR 2495.

References

Ashby, W. R. (1957): *An Introduction to Cybernetics*. Chapman & Hall, London..

Aunger, R. 2000 (ed.). *Darwinizing Culture: The Status of Memetics as a Scienc*e (New York: Oxford Univ. Press).

Aunger, R. 2002. *The Electric Meme: A New Theory of How We Think* (New York: Simon & Schuster).

Bertalanffy, L. von (1979): *General System Theory*, 6. Aufl., George Braziller, New York.

Black, D., 1995. The Epistemology of Pure Sociology. *Law and Social Inquiry 20*, pp. 856-858.

Basalla, G. 1988. *The Evolution of Technology* (Cambridge: Cambridge Univ. Press).

Blackmore, S. 1999. *The Meme Machine* (New York: Oxford Paperbacks).

Bouchard, T. J. (1997): „IQ Similarity in Twin Reared Apart“, in:Sternberg, R. J. und Grigorenko, E. L. (eds), *Intelligence, Heredity, and Environment*, Cambridge Univ. Press, Cambridge., 126–160.

Boyd, R. and Richerson, P. J. 1985. *Culture and the Evolutionary Process* (Chicago: Univ. of Chicago Press).

Boyd, R., Richerson, P. J. (2000): Memes: Universal Acid or a Better Mousetrap? In R. Aunger (ed.): *Darwinizing Culture: The Status of Memetics as a Science*. New York: Oxford Univ. Press, 143–162.

Bryant, J. M. 2004. „An Evolutionary Social Science? A Skeptic’s Brief, Theoretical and Substantive“, *Philosophy of the Social Sciences* 34/4, pp. 451–492.

Campbell, D. T. 1960. „Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes“, *Psychological Review* 67, pp. 380–400.

Cavalli-Sforza, L. und Feldman, M. W. 1973: "Models for Cultural Inheritance I: Group Mean and Within Group Variation", *Theoretical Population Biology* 4, pp. 42–55.

Cavalli-Sforza, L., Feldman, M. W. (1981), Cultural Transmission and Evolution. Princeton: Princeton University Press.

Cladière, N., Scott-Phillips, T.C., and Sperber, D. (2014). How Darwinian is Cultural Evolution? Phil. Trans. R. Soc. B 369, 20130368.

Cziko, G. A. (2001): „Universal Selection Theory and the Complementarity of Different Types of Blind Variation and Selective Retention“, in: Heyes, C. M. und Huber, L. (eds.): *The Evolution of Cognition*, MIT Press, Cambridge/Mass.), 15–34.

Dawkins, Richard 1976. *The Selfish Gene* (Oxford: Oxford Univ. Press, new ed. 1987).

Dawkins, R. (1982): The Extended Phenotype. Oxford: Oxford Univ. Press.

Dawkins, R. (1986): *The Blind Watchmaker*, *W. W. Norton & Company,* New York.

Dawkins, R. (1999): „Foreword“, in: Blackmore (1999).

De Jong, Gerdina 1994. "The Fitness of Fitness Concepts and the Description of Natural Selection," *Quarterly Review of Biology* 69/1, pp. 3-29.

Dennett, Daniel. C. 1995. *Darwin's Dangerous Idea* (New York: Simon & Schuster).

Diamond, J. (1998): Guns, Germs and Steel. London: Vintage.

Douven, Igor (2011): "Abduction", *Stanford Encyclopedia of Philosophy* (March 09, 2011, http://plato.stanford.edu/entries/abduction/)

Dray, W. (1957): Laws and Explanation in History. Oxford: Oxford Univ. Press.

Durham, W. H. (1991). *Coevolution: Genes, Culture and Human Diversity*, Stanford Univ. Press, . Stanford.

Edelman, G. M. (1987): *Neural Darwinism*, Basic Books, New York,

Eigen, M. (1971): „Self-Organization of Matter and Evolution of Biological Macro-Molecules“, *Naturwissenschaften* 58, 465–523.

Eigen, M., und Schuster, P. (1977): *The Hypercycle. A Principle of Natural Self-Organization. Part A*, *Naturwissenschaften* 64, 541–565.

Fisher, R. A. (1930): The Genetical Theory of Natural Selection, Clarendon, Oxford.

Futuyama, D. (1979): *Evolutionary Biology* (Sunderland/Mass: Sinauer Associates).

Gardner, M. 2000. "Kilroy was Here" (review of *The Meme Machine* by S. J. Blackmore), Los Angeles Times, March 5th.

Godfrey-Smith, Peter. 2009. *Darwinian Populations and Natural Selection* (Oxford: Oxford University Press).

Gould, S. J. (1991): Bully for Brontosaurus. New York: W. W. Norton.

Harman, Gilbert (1965): "The Inference to the Best Explanation", *Philosophical Review* 74, 88-95.

Henrich, Joseph, Boyd, Robert, and Richerson, Peter J. 2008. "Five Misunderstandings About Cultural Evolution". *Human Nature* 19, pp. 119-137.

Hodgson, G. M., Knudsen, T. (2006): Why We Need a Generalized Darwinism and Why Generalized Darwinism is Not Enough, Journal of Economic Behavior and Organization 61, 1-19.

Hodgson, Geoffrey M., and Knudsen, Thorbjorn. 2010. *Darwin's Conjecture: The Search for General Principles of Social and Economic Evolution* (Chicago: University Press of Chicago).

Hull, David L. (1982): „The Naked Meme“, in: H. C. Plotkin (eds.), *Learning, Development and Culture*, (Chichester: Wily), pp. 19-50.

Inglehart, R., 2003. *Human Values and Social Change: Findings from the Values Surveys.* Leiden: Brill Academic Publications.

Jablonka, Eva, and Lamb, Marion J. 2005. *Evolution in Four Dimensions* (Cambridge/Mass.: MIT Press)

Lewens, T., 2015. *Cultural Evolution - Conceptual Challanges.* Oxford: Oxford University Press

Lipton, Peter (2004): *Inference to the Best Explanation*, Routledge, London (2nd ed.).

Lorenz, K. (1941/42): Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie. Blätter für Deutsche Philosophie 15, 94–125.

Lumsden, C. J. and Wilson, E. O. (1981): *Genes, Mind, and Culture: The Coevolutionary Process*. Cambridge/Mass.: Harvard Univ. Press.

Maynard Smith, John 1974. "The Theory of Games and the Evolution of Animal Conflicts", *Journal of Theoretical Biology* 47, pp. 209–221.

Maynard Smith, J. (1982): Evolution and the Theory of Games. Cambridge: Cambridge Univ. Press.

Maynard Smith, J., Szathmáry, E. (1995): *The Major Transitions in Evolution*. Oxford: Oxford University Press.

Maynard Smith, J., 1986. Natural selection of culture?. *New York Review of Books, Nov. 6,* p. 33.

Mesarovic, M. und Takahara, Y. (1989): *Abstract Systems Theory*, Springer, Berlin.

Mills, Susan and Beatty, John. 1979. "The Propensity Interpretation of Fitness," *Philosophy of Science*, 46/2, pp. 263-86.

Millstein, R., 2010. The Concepts of Population and Metapopulation in Evolutionary Biology and Ecology. In: *Evolution Since Darwin: The First 150 Years.* Sunderland: Sinauer, pp. 61-68.

Müller, K. (1996): *Allgemeine Systemtheorie*, Westdeutscher Verlag, Opladen.

Mumford, S. 1998. *Dispositions*. Oxford: Oxford Univ. Press.

Nelson, R. R., Winter, S: G. (1982): An Evolutionary Theory of Economic Change. Cambridge, MA: Harvard University Press.

Niiniluoto, Ilkka: (2018): *Truth Seeking by Abduction*, Springer (Synthese Library), Cham.

Odling-Smee, J. (2010): Niche Inheritance. In: M. Pigliucci & G. Müller (eds.) Evolution - The Extended Synthesis. Cambridge/MA: MIT Press, 175-208.

Peirce, Charles S. (1878): "Deduction, Induction, and Hypothesis", in Peirce [CP], 2.619-2.644.

Perry, George, and Mace, Ruth. 2010. "The Lack of Acceptance of Evolutionary Approaches to Human Behaviour", *Journal of Evolutionary Psychology*, 8, pp. 105-125.

Popper, K., 1957. *The Poverty of Historicism.* London: Routledge.

Popper, K., 1979. *Objective Knowledge. An Evolutionary Approach.* Oxford: Clarendon Press.

Prior, E.W., Pargetter, R. and Jackson, F. (1982): "Three Theses about Dispositions", *American Philosophical Quarterly* 19, 251-7.

Quine, W.v. O. (1960): *Word and Object*; MIT Press, Massachusetts.

Ramsey, Grant, and De Block, Andreas. 2017. "Is Cultural Fitness Hopelessly Confused?," *British Journal for the Philosophy of Science* 68(2), pp. 305-328.

Rapaport, A. (1986): *General System Theory*, Abacus Press, Cambridge/Mass.

Reydon, Thomas A. C., and Scholz, Markus. 2015. "Searching for Darwinism in Generalized Darwinism," *British Journal for the Philosophy of Science* 66, pp. 561-589.

Ridley, Mark 1993. *Evolution* (Oxford: Blackwell Scientific Publications).

Salomon, S., 1992. *Prairie Patrimony: Family, Farming and Community in the Midwest.* Chapel Hill: Univ. of North Carolina Press.

Schatzki, T. R. (2002): "The Site of the Social: A Philosophical Account of the Constitution of Social Life and Change. University Park: The Pennsylvania State University Press.

Schurz, G. (2001): "What Is 'Normal'? An Evolution-Theoretic Foundation of Normic Laws and Their Relation to Statistical Normality", *Philosophy of Science*□28, 476-97.

Schurz, G. (2005): "Laws of Nature versus System Laws", in: Faye, J et al. (ed.): Nature's Principles, Springer, Berlin, 255-268

Schurz, G., 2007. Clash of Civilizations? An Evolution-Theoretic and Empirical Investigation of Huntington's Theses. In: C. Kanzian & R. E., eds. *Cultures. Conflict-Analysis-Diagnosis.* Frankfurt: Ontos Verlag, 277-294.

Schurz, G., 2008. Patterns of Abduction. *Synthese 164*, pp. 201-234.

Schurz, Gerhard 2011. *Evolution in Natur und Kultur* (Berlin: Spektrum Akademischer Verlag).

Schurz, Gerhard 2014. *Philosophy of Science: A Unified Approach* (New York: Routledge).

Schurz, G. (2021): "Evolution in Nature und Culture: Prospects and Problems of Generalized Evolution Theory", *American Philosophical Quarterly* 58(1), 2021, 95-110.

Skyrms, Brian 2014. *Social Dynamics* (Oxford: Oxford University Press).

Sober, Elliot 1993. *Philosophy of Biology* (Boulder: Westview Press).

Sober, Elliot und Wilson, David 1998. *Unto Others* (Cambridge/Mass.: Harvard Univ. Press).

Spencer, H. (1851): *Social Statics*, London.

Sperber, Daniel 2000. "An Objection to the Memetic Approach to Culture", in: Aunger (ed.), 163–173.

Steele, K., 2014. Choice Models. In: N. Cartwright & E. Montuschi, eds. *Philosophy of Social Science - A New Introduction.* Oxford: Oxford University Press, pp. 185- 207.

Vollmer, G. (1988): *Was können wir wissen? Bd. 1: Die Natur der Erkenntnis*, Hirzel, Stuttgart.

Voss, T., Abraham, M. (2000): Rational Choice Theory in Sociology: A Survey. In S. R. Quah, A. Sales (eds.): The International Handbook of Sociology. London: Sage, 50–83.

Ward, P. und Brownlee, D. (2000): *Rare Earth*, Springer, New York.

Weber, M. (1998): *Architektur der Synthese. Entstehung und Philosophie der modernen Evolutiostheorie*, de Gruyter, Berlin.Weibull, John 1995- *Evolutionary Game Theory* (Cambridge/Mass.: MIT Press)

Wilson, David S. 1999. "Flying over Uncharted Territory: A Review of *The Meme Machine* by S. Blackmore", *Science* 285, p. 206.

Wilson, Edward O. (1975. *Sociobiology. The New Synthesis* Cambridge/Mass.: Harvard Univ. Press).

Chapter 11

Human Social Evolution via Four Coevolutionary Levels

Theodore Koditschek[i]

Abstract

Social evolutionary theories have been roundly discredited since their original racist and sexist formulations during the nineteenth century. In light of recent advances in evolutionary biology and anthropology, is it now possible to reconstruct alternative paradigms of social evolution in scientifically defensible terms? This chapter proposes a novel framework along these lines, involving four analytically distinct but empirically nested levels (and logics) of evolution. Above the biological level, with its logic of natural selection, it follows recent scholarship on the cultural level (with a non-reductive logic of cultural selection). But then it adds two further non-reductive and emergent levels: a political level, grounded in a 'logic of domination', followed by an economic level, driven by a 'logic of capitalist competition' to track human evolutionary history up to the present day. The chapter considers how far these additional levels/logics can draw on Darwinian mechanisms of selection *between* variational units, and where it must shift to an examination of transformational processes *within* the units themselves. It explores the role of agency in the evolution of human societies, and how this relates to the more systemic selectional and variational processes. It concludes with an assessment of the intellectual benefits that such retrodictive accounts of *longue durée* evolutionary history might bring to the social sciences.

---

[i] University of Missouri, Columbia. Needless to say, any errors in my own understanding of this material should be attributed to me alone.

It is well known that most humanists and mainstream social scientists are extremely skeptical (if not downright hostile) to the proposition that evolutionary theories have any real value in explaining the course of human affairs. The reasons for this hostility are basically twofold: 1) Political aversion to an intellectual tradition that was deeply inflected with racism and ethnocentrism in its original nineteenth century formulations, and which had been used (right up through the 1930s and 1940s) to justify sexism, colonialism, and genocide. 2) Professional disinterest in the ways in which scientific advances over the last century in paleontology, archaeology, radiometry, molecular biology, population genetics, epigenetics, human ecology, anthropology, psychology, behavioral economics, game theory, agent-based modelling, and computer simulation have altered the foundations of evolutionary thinking -- making it possible to develop theoretical approaches that will underpin cultural and social evolutionary theories of a very different kind (Degler, 1991, Segerstråle, 2000, Laland and Brown, 2002).

To dispel contemporary fears of a return to toxic and outmoded versions of ideological social Darwinism will not be easy. To rectify professional disinterest about potentially relevant scientific advances in other fields will be tough in a different way. Taken together, getting evolutionary perspectives back on the map of mainstream social scientists will be a task of herculean proportions. For both sound and regrettable reasons, I do not think it is likely to succeed anytime soon. What is possible right now is for sympathetic interdisciplinary practitioners to take the lead in advancing incremental ideas, models, and proposals for how evolutionary social science might be done responsibly (and why it is worth doing). The skeptics will object, and these objections will have to be addressed. As ideas are clarified and as models and proposals are improved, extended discussion will assess the validity and utility of the project. This is how science works. As sharper, more nuanced, and more persuasive heuristic frameworks are placed on the table, evolutionary social science might be taken more seriously by a wider audience.

In this paper, I hope to contribute to this enterprise by building on what I regard as the most generative breakthrough in the effort to construct evolutionary theories that are grounded in biology but are in no way dependent on biological reductionism. These are the various theories of gene/culture coevolution, which have been proposed by a small but significant cohort of anthropologists and their interdisciplinary allies[1] These theories are not without their problems and limitations. Moreover, from my own perspective as a social historian, they are woefully inadequate in explaining the main lines of human social evolutionary development that have played out over the past five thousand years. Nevertheless, they offer a promising start, and a large part of what they promise is a fruitful model for integrating separate strands of coevolution – biological, cultural, political, and socio-economic -- that are constantly interacting, as they each operate at a different level of intensity.

My approach therefore builds on the gene/culture coevolutionary tradition, albeit in a manner that incorporates some inescapable complications. In an effort to address these complications with utmost parsimony, I have added two further emergent levels to the existing coevolutionary standpoint. I believe this move will go a considerable distance in accommodating those features of human social evolutionary development that the existing versions of the theory ignore. In my account, adaptive genes are selected to produce creatures capable of selecting adaptive cultural variants. But these self-modifying cultures must then 'learn' to interact with political entities that have emerged with the power to shape the cultures they encompass. Because these polities find themselves in competition with other polities, the most successful variants will be selected as prototypes for polities of the future, which will then be shaped by a fourth additive level of capitalist evolution that is driven by a different set of optimizing competitive encounters between more or less adaptationally adept agents, economic enterprises, and capitalist states.

Taken together, these evolutionary strands all exert an ensemble of selective cross-pressures on variations or mutations that appear on multiple levels (sometimes to reinforce one another, sometimes in contrary directions). The result is to shape a complex course of social development, which may seem on the surface to be random and fortuitous but that may reveal

---

[1] For my understanding of progress in developing the gene/culture coevolutionary theoretical framework I have drawn primarily in the following sources: Cavalli-Sforza, and Feldman, 1981; Durham,1991; Sperber,1996; Richerson, and Boyd, 2005; Bowles and Gintis, 2011; Mesoudi, 2011; Lewens, 2015; Henrich, 2016; and Laland, 2017. I have also benefitted enormously from my many conversations with Karthik Panchanthan over many years. As a master of this literature and a superb practitioner of its craft, Karthik has been an invaluable resource in guiding me through the challenging issues raised by this body of work. I also benefitted greatly from auditing his graduate seminar, and from the many Evolutionary Social Science lecturers that he has brought to present papers at the

subtle patterns of cross-cutting determination underneath. One way to understand these divergent (but interacting) modes of evolution would be as a succession of nesting dolls that constantly operate inside (and upon) one another. Like the dolls, the levels can be analytically unpacked to understand their separate developmental logics. Yet, after this is done, they can also be synthetically re-assembled, so that their complex, mutually constitutive coevolution can be understood in its totality. Of course, since nested dolls are not dynamically interacting, the simile with my levels of coevolution is inexact. But they may offer a useful visual metaphor to begin the discussion.

In 2019, I published a paper (Koditschek, 2019) where I laid out, in broad stokes, how such a four-level social evolutionary theory might work[2]: How it might encompass (at least in a very coarse-grained way) many of the major features of social and historical development that human cultures, societies, polities, and economies have undergone, especially as history has recorded them over the past five-thousand years. In this paper, the two additional levels (atop gene/culture coevolution) are characterized as 1) A 'logic of domination' which first emerged c. five thousand years ago, and 2) A 'logic of capitalism', which emerged atop the gene/culture/domination ensemble, some two hundred years ago. By treating these new coevolutionary levels as emergent properties, I proposed a coevolutionary model of human historical evolution that is in some respects analogous to the Maynard Smith – Szathmary model of the major transitions in the evolution of life. (Maynard-Smith, Szathmary, 1995, 1998)

In the present paper, I wish to build on the claims of my previously described model, but also to point out some of the problems and challenges that emerge from a closer look. The nested dolls (i.e. the ascending logics of coevolution) do fit closely inside one another, but they do not all fit together in exactly the same way. As the advocates of gene/culture coevolution

---

[2] Although this article is the fruit of several decades of my own reflections as an historian, the citations indicate many debts to previous thinkers and researchers. Special mention must go to the sociologist W.G.A. Runciman (2009), who almost singlehandedly pressed the case for the value of a quasi-Darwinian (but fundamentally non-biological approach to social evolution, at a time when few others in his (or allied) disciplines were prepared to take this endeavor seriously. In my own essay (2019) I critically assess Runciman's contribution. In footnotes 1, 2, 5, and 9 of that essay, I have laid out some of the reasons why my proposed approach to social evolution avoids the political and epistemological defects that have been charged against older evolutionary approaches to the social sciences. I hope to make this case more fully on some future occasion.

have demonstrated, the logic of cultural selection that produces change over time in the distribution of cultural variants is in some respects analogous to the logic of natural selection that regulates the biological evolution of species and varieties. But in other respects, the two are not identical. Correspondingly the logics of domination and of capitalism in my model have the effect of altering the selective landscape, not merely by adding two further levels of selection, but by changing the determination of who/what selects in these two 'higher' domains.[3] The result is that the simple Darwinian trifecta of variation, selection and inheritance becomes complicated with the increasing force that is exerted by selection and mutation in these 'higher' domains. Evolutionary accounts, which may be 'variational' when viewed primarily at the lower levels, require the increasing incorporation of 'transformational' elements as one moves up the ladder of selection/mutation domains. As biologists are now discovering in their studies of organic evolution, the rigid dichotomy between evolutionary (phylogenetic) and developmental (ontogenetic) processes becomes less-clear cut as one moves to richer accounts (and higher-level examination) of societal evolution. As we extend evolutionary understanding beyond the simplest principles of natural selection, some modification of foundational concepts will also be required. In this approach there is partial consilience between the four levels of evolutionary explanation, but they do not lock together in a strictly reductionist way. Each level, as it is added, connects with the one 'below' it by unleashing emergent properties into the human experience. As these properties open novel vistas, they not only shift the course of species evolution, but also change what human evolution means.

---

[3] When speaking of levels of coevolution, I am using 'higher' and 'lower' in a value-neutral sense, simply to indicate that they are chronologically overlaid atop one another in terms of their order of emergence. In the experience of human subjects, at any moment in history, the levels may not be perceived as clearly delineated, since the separate logics constantly interpenetrate and interact. Nevertheless, it is generally the. case that human subjects experience selective pressure to adapt with greater intensity from the highest level/logic, whereas lower levels/logics usually exert their adaptive pressures with relatively diminished force. The exception comes when adaptation to higher level selective pressures bring lower-level institutions to the point of fundamental jeopardy. Under these circumstances, the need to sustain the lower-level institutions will force a re-calibration of the higher-level pressures and practices so that they will be consistent with the basic maintenance of the polity, culture, or human body/psyche. Although this caveat may strike the reader as somewhat abstract or obscure right now, I will try to re-articulate it more concretely in a note at a later place in this paper, when the point should become clearer.

11.1. How to Explain the Human World?

Kevin Laland begins his account of gene/culture coevolutionary theory by contrasting what Darwin saw from his Down House garden with Laland's own view out the window of his university office. Where Darwin apprehended "an entangled bank, clothed with plants of many kinds, with birds singing on the bushes, with insects flitting about, and with worms crawling through the damp earth." Laland saw "stone buildings, roofs, chimneys, … telegraph poles and electricity pylons." He wondered "can evolutionary biology explain the existence of chimneys and cars, and electric pylons in as convincing a fashion as it does the natural world?" Laland concludes that biological evolution alone cannot explain this outcome, but by adding culture, such understanding can be gained. First, it required the biological evolution of a creature capable of complex culture, but then the further development of cultural forms that impacted on the process of biological evolution, creating "a whirlpool of coevolutionary feedbacks in which culture played a vital role." (Laland, 2017, p. 3)

Surely, however, there is a fundamental difference between Darwin's concluding, and Laland's opening assertions. We assent to the "grandeur in this [Darwin's] view of life," because of the way his theory of natural selection can actually account for "these elaborately constructed forms, so different from one another, so dependent on another" as they have been brought into being by the "laws [he enunciated] acting all around us." (Darwin, [1859], 2001, p. 174). By contrast Laland's conjuring of cars, electric pylons, and chimneys" from a process of selection by "culture" is a hand-waving gesture that tells us nothing concretely about how these objects have become entangled in our lifestyle, or the processes by which they have been brought into being as central features of our physical and social world. After all, human culture in some form has been present for at least a million years but, until the last five thousand years, it produced nothing more than lithic technology, small, informal social organization, and the first stirrings of plant and animal domestication. It is a long way from this to cars and electric pylons. Vague invocations of "cultural selection" alone, will not bring us from the stone age to the age of electronics.

To explain the pylons and the cars, etc., it will be necessary to introduce a whole host of institutions, power relationships, and technological processes that "culture" alone will never account for: corporations, property law, market exchanges, production facilities, divisions of labor, class hierarchies, gendered norms, bureaucratic structures, governments, (both national and local), colleges of technical training, institutes of scientific research, and networks for the diffusion and implementation of discoveries – just to cite a few examples. There is no reason, in principle, why these things cannot be accounted for in an evolutionary manner, since every one of these institutions and practices has also evolved. It will simply require an explanation that operates more complexly than gene/culture evolution. Such an explanation, which I propose to handle through two additional logics, must tell us 1) how institutions and practices of this type first came into existence (at a time when they had not existed before) and 2) how they then subsequently evolved in coordination/competition with one another, as well as the ways they were navigated and directed by human agents who had themselves evolved in a set of pre-existing (but not unchanging) cultural ensembles, inhabiting bodies that had been shaped by natural selection, which also continues to subsist as a deep, background force. My goal in this paper, as in its predecessor, is to produce the most conceptually pared-down framework in which such a (four-layered) coevolutionary theory can be laid out. (Koditschek, 2019)

Let us consider how the four nested logics that produced this constellation can both bear a striking family resemblance, and yet differ in consequential ways. We may begin with the two logics whose relationship has already been explored by Laland and his colleagues in the gene/culture genre of 'dual 'inheritance' scholarship. There is indeed nothing completely new in the idea of gene-culture coevolution, since a very fuzzy version of it has been circulating since Darwin's day. The notion was then fuzzy because the very distinction between nature and culture remained unclear throughout the nineteenth century. Before Weismann's experiments on mice (and for a considerable time thereafter) there was a widespread belief that acquired characteristics could be directly inherited. As a result, it was easy to envision cultural attributes and innovations being passed down through bloodlines, while the "blood" of individuals from different backgrounds could be seen as differentially reflecting the cultures which produced the individuals in question. (Bowler, 1983, Koditschek, forthcoming). While Weismannism ended this racist, sexist confusion, the radical separation of nature from culture that became normative in the twentieth century engendered a different set of conceptual limitations. The now separated studies of culture and of nature grew so distant from one another that social and biological sciences lost the ability to communicate with one another, developing separate languages, methodologies, operating assumptions, as well as a tacit agreement to each confine themselves to separate (and completely disjunctive) research spheres. When sociobiology threatened this gentleman's agreement in the 1970s, it caused understandable alarm in the community of social scientists, which feared that a new campaign of biological reductionism was poised to re-introduce the racist and sexist assaults on social science that had deformed the thinking of the nineteenth and early twentieth centuries. (Sahlins, 1976, Montague, ed., 1980, Degler, 1991, Segerstråle, 2000).

Real progress came however, during the 1980s and 1990s with the recognition on the part of four pioneering interdisciplinary scholars, Luigi Luca Cavalli Sforza, Marc Feldman, Robert Boyd, and Peter Richerson, that many of the same statistical techniques, which had been developed in population genetics to understand changing distribution of particular gene alleles, could also be used to understand the evolution of cultures. This could be done by tracing the differential spread of atomic cultural elements (variously named 'memes', or 'cultural units') transmitted solely by (non-DNA) informational methods in a non-reductive, non-biological way. Since this transmission of acquired cultural beliefs and practices could be accomplished by teaching, copying, indoctrination, or other 'Lamarckian' means, selective forces could operate more swiftly and purposively than at the genetic level, where the introduction of new alleles occurs only sporadically, by random mutation. (For references see note 1). When these shifts in culture were sufficiently rapid and comprehensive, such changed cultural behavior at the collective level could have feedback effects on the biological transmission of genes, through the mechanisms of group selection that rarely operated on other species, where culturally driven behavior was absent or weak. (Sober and Wilson, 1998, Richerson and Boyd, 2005, Bowles and Gintis, 2011).

Not all the problems of analogizing population genetics methodology for the study of cultural evolution have been satisfactorily ironed out. Can culture really be disaggregated into elementary particles? Are these to be understood as ideas, practices, bits of know-how, or beliefs? How are these cultural particles transmitted or learned? Do mutations in ideas, practices, bits of know-how, or beliefs appear randomly? Are they transmitted in a directed and/or biased way? To what extent are cultures better understood as integrated ensembles in which the parts cannot be separated from the whole? We need not pursue these questions in this venue. It is sufficient to note that at least some of them have been creatively addressed by the practitioners of gene/culture coevolution (Sperber, 1996, Mesoudi, 2011, Lewens, 2015). What is more telling is that, among mainstream social scientists and historians, this genre of scholarship remains almost entirely unknown. This is partly the result of disciplinary silos, partly a result of allergy to the very word 'evolution', and partly to distrust of elaborate mathematical techniques. But it is also a product of the fact that the currently well-documented cases that can be uniquely explained by gene/culture coevolution are relatively trivial and few. The best known is the spread of a gene for adult lactose tolerance among populations whose ancestors adopted dairying agriculture over the past few thousand years. Here is one thoroughly documented case in which a drastic change in culture (the shift from hunting and foraging to the domestication of dairy animals) has led to a gradual change in the biology of digestion. Other small biological changes, having to do with skin color, heat tolerance, and salt retention may well have been precipitated by cultural change in this way. (Durham, 1996, p. 226-255, Henrich, 2017, p. 54-96, Lewens, 2015, 89-93).

The hypotheses of current practitioners of gene/culture coevolution analysis however, go far beyond these relatively minor effects. The work of recent synthesizers such as Kevin Laland, and Joseph Henrich have speculatively (albeit quite plausibly) offered far reaching suggestions that many of the most unique (and hitherto evolutionarily inexplicable) characteristics of our species were acquired in this way. Our intelligence, our capacity for abstraction, our self (and other) consciousness, our capacity for language, our behavioral plasticity, our tendency towards cooperation, our forms of marriage, kinship, and childrearing, our propensity towards tribalism, and our altruistic impulses towards non-kin are all posited to be the products of the same gene/culture coevolution in these recent accounts. Through an increasingly sophisticated capacity for culture, these accounts hypothesize, our species (and its immediate hominid predecessors) have evolved a host of unique biological features over a relatively short time period. Such rapid species change in anatomy, physiology, and behavior would be difficult or impossible to explain by the normal operation of natural selection on DNA alone. According to these accounts, significant gene/culture coevolution began for our ancestors perhaps some 1.5 million years ago, proceeding sedately and cumulatively at first. But then this coevolution speeded up about 500,000 years ago with more sophisticated lithic technology, enhanced social networking, and control of fire (Henrich, 2016, p. 280-332, Laland, 2017, p. 175-182).

Sexual dimorphism steadily diminished, hunting, gathering and toolmaking gradually reshaped limbs, fingers, muscles and neural networks, as jaws atrophied and skulls expanded. Cooking transformed the physiology of digestion, as stomachs shrank to provide more resources for brains. With the advent of modern *Homo sapiens*, (c.150,000 years ago) culturally driven biological change became a runaway process. Language, complex communication, symbolic representation, and collective decision making all became self-sustaining, and mutually reinforcing through a process that culminated in the recognizably contemporary human beings who have come to dominate the planet over the past c. 40,000 years. (Klein, 1999, Henrich, 2016. See also Chapais, 2008, Hrdy, 2009). Presumably, this process of gene/culture coevolution is still going on, albeit as a pace that is now dwarfed by the even more rapid cultural/political coevolution that has emerged on top of it, as well as the yet more rapid cultural/political/economic coevolution that has become characteristic of the most recent capitalist age.

11.2. The Logic of Domination

My hypothesis is that the two additional logics (and feedbacks) of coevolution that my paper posits also began comparatively slowly, but gradually sped-up to become runaway, autocatalytic processes that redirected further additive vectors to human evolution, once they reached the point of irreversibility. The result was to create a more complex evolutionary landscape in which the need to adapt first to cultural, then to geo-political, and finally to economic market pressures successively dwarfed each of these pre-existing adaptive pressures, without ever making them go away entirely. Let me explain briefly how this snowballing (or doll nesting) dynamic seems to have worked.

By the 'logic of domination', I mean the emergence of state (or quasi-state) coercive structures designed to constrain those within their boundaries by force and legal institutions. As time went on, these were increasingly supplemented by ties of economic interdependence and ideological instruments for inducing consent. In the final instance, domination was secured by the public state authorities, but it was increasingly also reinforced by other forms of domination, such as class hierarchy, slavery, and patriarchy in the private domestic sphere. The first city-states states emerged c. 3500 BCE in Mesopotamia, as well as semi-independently along the Egyptian Nile (c. 3000 BCE), and in the Indus valley (c. 2500 BCE), and then fully independently in the Guatemalan Highlands by 1000 BCE. At first, these city-states were small and unsteady, but the initial iterations spawned imitators, which quickly became competitors in an arms race that was both figurative and literal: The imperatives of winning engendered feedback cycles in which the ability to mobilize internal resources, labor, population and compliance fed further advancements in bureaucratic control and military technology, (as well as *vice versa*): Those polities that proceeded through these cycles most quickly and efficiently came out as winners in an inter-state rivalry and, as they won, the velocity of the cycles increased (Koditschek, 2019, McNeil, 1982).

Nevertheless, the earliest states were extremely fragile constructions, prone to epidemic disease and ecological crisis, in which control had to be exerted in draconian (and often self-defeating) fashion, because it was relatively easy for subjects to defect by simply escaping to the hinterland beyond. Before long, successful states were the ones in which a certain symbiosis was achieved between humans, their domesticates, and parasitic micro-organisms. At that point larger swaths of adjacent territory could be acquired and controlled by the state authorities, while powers of naked coercion could be supplemented by subtler mechanisms of ideological consent (usually managed by increasingly organized priesthoods in state religions). And so, the most successful city-states evolved into territorial states with kingships, and these evolved in turn into regional empires, in which multiple cultures with differing languages (and sometimes different religions) had to be accommodated. Epidemic disease, and environmental disaster continued to erupt as periodic catastrophes, for example, in the 'Dark Ages' that descended on the entire Mediterranean region, sometime c. 1100 BCE (Cline, 2014, Scott, 2017)

During these periods of political devolution, structures of domination were loosened, freedom was regained for the lower classes, and yet, all too often, anarchy spread. New patterns of local control then emerged, with new institutions of warlordism as a response to chaos. And so the cycle (of centralized state and class domination) began (at accelerated pace) all over again. As the local (and then regional) states evolved into complex polities, those who resisted were pushed increasingly further outside onto the margins, and onto the frontier marchlands where they became discursively constructed from the center as 'savages' and 'barbarians' i.e. those who had successfully resisted central control (McNeill, 1991, Turchin, 2003). While the savages (hunters and foragers) could be completely marginalized (and, where necessary, exterminated), the barbarians, with their own complex economies of pastoralism, often became formidable military antagonists, who engaged in raiding and trading with the imperial-state centers. Occasionally they invaded to replace the established dynasties and aristocracies with entirely new personnel of central control. In fact, when viewed in the wide-angled lens of *longue durée* history (that is, the history of centuries, and ultimately millennia), centralized states and their barbarian antagonists on marchland peripheries are best seen as symbionts – shuttling back and forth in a cyclically re-balancing structural order, but also keeping that order flexible and porous to the exchange of goods and the introduction of people and ideas from beyond the borders of the centralized imperial-state zones (Braudel, 1966, Golden, 2011).

One way to trace the advance of this general logic of domination as a force around the globe would be to calculate the area (and estimate the population) within state boundaries (or at least under significant centralized legal, bureaucratic taxation, and military control) at various dates in history. This would enable one to calculate changes in this variable as a proportion of the world's territory/population as a whole. Such calculations would be at best tentative, approximate, and are in any case, beyond my ability to offer right now (See Maddison, 2006, 2007, for some numerical building blocks). I would however predict an erratic (albeit regionally cyclical) pattern, as state-power went through several long-term cycles of regional expansion and partial contraction (as epochs of advancing 'civilization' were followed by partial relapses into 'dark ages' of barbarism), after which the cycle would begin (with altered central state loci) at a higher civilizational starting point. On a global scale, the peaks and troughs would likely cancel one another out, except in those rare instances when epidemic disease or ecological catastrophe had a uniform effect over the entire planet (Turchin, 2003, Parker, 2013).

Over the very long run, however (i.e. over the entire span of the last five thousand years) the data would surely show that the logic of domination was proceeding at an accelerating pace. The pace would be (by later standards) slow, and regionally dispersed, up to about 1000 BCE. Thereafter it would begin to pick up speed and spread over the Eurasian trans-continental mass. It was then that Eurasia entered the 'Axial Age' of the classical ancient civilizations, which mandarin intellectuals like Karl Jaspers and Shmuel Eisenstein have celebrated as the birth of civilization, while maverick scholars like James Scott and David Graeber have recently characterized as ages of disaster for lower class plebeians, debtors, slaves, peasants, and the poor (Eisenstein, ed., 1983, Jaspers, 2021, Scott, 2017, Graeber, 2011). The final 'dark age', which then descended after the decline of the Roman Empire in the west, paved the way, some six centuries later, for the (less bureaucratically centered) resurgence of the 'high middle ages'. After a brief fourteenth and fifteenth century trough, this was followed by a now-irreversible 'early modern' upswing in state-building between the sixteenth and eighteenth centuries. Spreading from its initial locus in the European west, this new surge of state re-building was inextricably connected with the European conquest of America and ascendancy over a series of hitherto autonomous Asian empires. With its heightened bureaucratic, legal, military, and technological superiority, these tightly centered early modern European polities were able to create new trans-oceanic empires that spread their distinctively evolving nation-state forms everywhere around the world. (Aston, ed., 1965, Anderson, 1978, 1979, Wallerstein, 1976, 1980, 1989).

11.3. The Logic of Capitalism

As much recent work has demonstrated, the ability of westerners to achieve this position of full global dominance was partly the result of a series of accidents that gave them a temporary ecological, epidemic, and demographic advantage over other peoples in what became a set of transcontinental Atlantic, South Asian, and East Asian exchanges of peoples, domesticates, slaves, microorganisms, viruses and goods (Crosby, 1972, 1986, Blaut, 1993, Pomeranz, 2000, Parsathrathi, 2011). However, I would maintain that it was fundamentally fueled by the emergence and maturation of a radically new logic of global societal interconnection: the logic of capitalism. Of course, trade and commerce had been marginal features of even the oldest human hunting and forging communities. During the early eras of state-building they were supplemented by money, debt, and taxation, as trade and commerce began to assume increasingly trans-political forms. Nevertheless, up through the European middle-ages, mercantile and profit-making activity had always been contained (and constrained) within the interstices of the state (or feudal) political forms (Polanyi, 1944, Pirenne, 1952, Hilton, ed., 1978)

During the early modern period this pattern continued, but the quantitative expansion of trade and commerce reached a tipping point. This was the result of a shift in the power balance between the *profit-making activity* of entrepreneurial capitalists and the *political* muscle of the incipient western European imperial nation-states. Relying on revenue from commercial expansion and plantation colonies for their superior bureaucratic and military power, these states authorized a new level of economic initiative based on the mercantilist premise of quasi-autonomy. Eventually, this enabled capitalism to break out of its mercantilist fetters and to transcend the political chrysalis in which it had gestated. From that point on it was free to disclose its own novel and distinctive logic of social organization, and to pass its own threshold of irreversibility. During the nineteenth century this new logic of capitalism became the cutting edge of a system for coordinating and reorganizing a single mode of social organization, first in western Europe, and then everywhere around the globe (Findlay and O'Rourke, 2007, Fox-Genovese and Genovese, 1983, Abernethy, 2000).[4]

Like the rise of the state before it through the logic of domination, the rise of capitalism through a logic of increasingly globalized market exchange began as a 'singularity'. It began in the west, not through any intrinsic western superiority, but merely as the result of a series of accidents that happened to favor western power in this particular time and place. Even in the west, capitalism's initial development gathered steam slowly, and was long constrained by extraneous forces. However, sometime between the sixteenth and eighteenth centuries, it became essential to the continued development of the most advanced pre-existing western states. There, in Britain, the Netherlands (and to a lesser extent France) revolutionary liberalism overthrew aristocratic and corporatist old regimes. Thus, the way was opened for the liberal individualism of the new dispensation to become the dominant (eventually the inescapable) mode of social organization for human communities everywhere around the world (Hobsbawm, 1962, 1968, DeVries, 1976). As with the logic of domination, it should be possible to compile a quantitative index of the rise and spread of capitalism as a globalizing force. Today, we can safely use world and national GDP as a rough index of this process. However, before the nineteenth century quantitative estimates of GDP are crude, regionally uneven, and ultimately meaningless as a tool for comprehensively measuring overall economic activity, most of which

---

[4] Beyond these (and other) empirical sources, my claim that capitalism inaugurates a new level rather than just intensification of pre-existing market exchange is some basic argument grounded in classical Marxist theory. The central assertion here is that proletarianization of the working class (turning the mass of ordinary producers into wage workers rather than slaves, serfs, independent peasants, artisans, hunters, foragers, etc.) sparks a fundamental change in entire social organization. Not only does it turn the mass of the population into hired laborers, it also makes them compulsory commodity consumers, rational choice actors, and ultimately (when they collectively demand it) citizens as well (Marx and Engels, 1948, Marx, 1954). It transforms the family, breaking it up as a unit of production, leading to the emancipation/domestication of women, which in turn underwrites public education, the social service professions, Foucauldian disciplinization (Foucault, 1979), and the mass conscription state. On another level, it turns productive property from mere legally privileged ownership into compulsorily market exposed (i.e. de facto collectivized) capital, so that competition between capitalists now necessitates constant transformation of production processes, in which increasing social wealth is the consequence, not the cause. Yet, I would go beyond these Marxist arguments to assert that the autocatalytic expansion of capitalism into a globally dominant system constitutes the point when it becomes fully instantiated as a new dominant *logic*. My claim here is that it is the *irreversibility* of these interlinked transformations that ultimately warrants treating capitalism as not merely a new social system, but a new *level* and a new *logic* of evolutionary change. Once people, cultures and states have passed this threshold it is simply impossible to go back to the old ways, absent the collapse of human civilization.

still occurred in subsistence and barter economies outside the monied sphere of market exchange (Maddison, 2006).

A better measure is one that captures the moment, and documents the dimensions, of the passage of capitalist logic to the point of irreversibility. This was the point, as Marx and Engels argued 173 years ago, when the superior manufacturing and factory methods of commodity production that emerged under its aegis enabled western capitalists to "batter down all Chinese walls" of traditional commodity production and exchange. With the help of liberal free trade ideology and its imperial state domination of peripheral regions, western capital was able to crush traditional methods of handicraft production and to compel artisans, peasants and laborers in more and more portions of the globe to enter the labor market as proletarians in commodity production and as consumers for capitalistically fabricated goods (Marx & Engels, [1848], 1948).

Since this is my own area of research specialization, I can offer one quantitative index of this process that exposes the moment of transition when the first (British) capitalist industrial revolution unleashed its transformative potential on the then still precapitalist economy of most of the globe. As I see it, the crux of the cutting edge of this industrial capitalist revolution can be identified by the confluence of three quite precise measures. 1) When the supply and demand of any given commodity suddenly expands by explosive proportions. 2) When the cost per unit declines equally dramatically (usually, but not always, as the result of labor efficiencies brought about through mechanization), and 3) When these *local* transformations in the mode of producing one specific commodity feed-back into a self-sustaining and self-reinforcing process that autocatalytically spreads the same dynamic logic to other commodities in the same and eventually in other production spheres (Rostow, ed., 1963). So, in Britain, between 1770 and 1850 the output of cotton yarn increased 150-fold, which was accompanied by a decline in price per unit of 90%. This corresponded to the mechanization of spinning in the late eighteenth and early nineteenth century. The mechanization of weaving and other finishing processes during the 1820-50 period led to a further price decline in cotton cloth to 5% of its former value, accompanied by a 160-fold expansion of both supply and demand. By 1850 then, industrialization had spread from cotton to virtually all the textile industries. When combined with a less dramatic expansion in iron and coal production (55x for iron, 6x for coal) and a decline in iron prices to approximately 20% of their former value, the path was opened for a fully self-sustaining autocatalytic drive. Industrializing techniques now spread throughout the entire economy, as coal, steam and iron were combined to produce the railroad, the steamship, with the world's first fossil fuel energy regime. (Calculated from data in Baines, 1835, Mitchell, 1962, Deane and Cole, 1962, Von Tunzelman, 1978, Koditschek, 1990).

Although my own quantitative analysis drops off around 1850, the autocatalytic dynamic of capitalist industrial revolution and expansion did not stop there. Over the next few decades, it spread to the production of steel, chemicals, machine tools, and pharmaceuticals, not only in Britain, but also in Germany and the United States. During the twentieth century the same dynamic further metastasized to encompass agriculture, domestic appliances, automobiles, electronics, electrification, aerospace, and mass entertainment via film, phonograph, radio, television, and print. If global measures of total factor productivity are to be taken as an index, the data shows that by the 1950s and 1960s industrial capitalist production had become dominant throughout the entire world. Since then, of course we have seen a further wave of intensification in the global capitalist revolution with the advent of digitization, the internet, social media, and biotechnology, as well as the spread of post-industrial economies driven by high finance and multi-national corporations, which have further transformed both the developed and the developing world (Maddison, 2001, Gordon, 2016).

11.4. What Does this Add to What We Already Know?

By tracing the vicissitudes of the logic of domination, and then by piling an account of the logic of capitalism on top of it, my evolutionary argument may not seem to add much in the way of straightforward facts to what we already know. Conventional historians and sociologists will half-acknowledge the validity of these sequences – shorn of their evolutionary pretensions – on the rare occasions when they look beyond their narrow field specializations to survey human history as a whole. In recent years, a few sweeping 'bird's eye' narratives of the human experience have been written in this spirit; Jared Diamond's *Guns, Germs, and Steel* (1997), J.R. and W.H. McNeil's *The Human Web* (2003), Ian Morris's *Measure of Civilization*, Yuval Harari's *Sapiens* (2015), or, even more expansively, Cynthia Brown's *Big History: From the Big Bang to the Present* 2007), or David Christian's *Maps of Time: An Introduction to Big History* (2014). These ultra-*longue durée* histories make valuable contributions to the education of the general public, but they are quite different from what I am trying to do. My aim is not to produce another grand narrative of the life-cycle of our species. It is rather to sketch out a series of theoretically linked lines of explanation in which practicing historians and historical social scientists can locate their more granular findings within a common overarching evolutionary frame. If these linked lines of explanation are taken up, it may become useful to craft periodic syntheses of the state of provisional play across a series of evolutionary disciplines, taking relevant research into account. But the lines of explanatory dynamics themselves should be judged on the basis of 1) their internal theoretical consistency and capacity to make sense of empirical results, and 2) their external theoretical consilience in building upon one another in a broadly defined evolutionary frame.

Of course, the postulation of broad theoretical frameworks to explain the details of human historical development is also nothing new. From Thomas Hobbes, Adam Smith and Karl Marx, to Herbert Spencer, Max Weber, and Arnold Toynbee, various grand theories of history have been proposed from time to time. In the past seventy years the most robust of these these theories have been refined by sympathetic academics into practice-ready paradigm proposals that have served as stimuli for my own work for several decades.[5] (Kuhn, 1962). So, for example, the reader may already have noticed that my 'logic of domination' bears some resemblance to theories of geopolitical determinism that have been derived by modern scholars from the Hobbesian tradition. Correspondingly, my 'logic of capitalism' draws from two paradigms, Marxism and Modernization – often seen as antithetical, which place the rise of markets and production transformations at the center of a general historical dynamic. Where Modernization identifies markets and the division of labor as the *drivers* of social transformation, Marxism builds on (and inverts) these insights by focusing on the *contradictions* left by the Modernization account.

When adopted separately, however, each of these theoretical paradigms is incompatible with the others. My strategy, by contrast, has been to render them interdependent by placing

---

[5] I take my basic understanding of scientific 'paradigm from Kuhn, 1962). The application of Kuhnian concepts to various social sciences (including history) is examined in Hollinger (1973), Perry (1977), and Eckberg and Hill, (1979), although I have not found any of these interesting and informed accounts to be entirely suited to my purposes.

them together (with the necessary modifications) in an overarching evolutionary frame.[6] The success of this move depends on the extent to which I have persuasively separated each of these theories from the respective teleologies of 'progress' that marred their original formulation. It further depends on the extent to which my approach is able to strip each one of the dogmatic foundationalism that would alternately privilege a single dimension of the human experience (e.g. the political, the economic, or the social) as the driving force behind everything else. It is to solve these problems that I have introduced my notion of nested logical levels, each with a set of different emergent properties. To acknowledge these differing logics, without allowing any one to dominate, I have proposed a strategy that would begin by analyzing them separately, with the aim of ultimately seeing them recombined and studied in terms of their interaction: The logic of natural selection will appear as modified by the logic of cultural selection, which will in turn be further modified by the logic of domination, and then by the logic of capitalism, at which point, each and all will have been substantially redirected and changed.

If my work has a unique and original claim, it is that this comprehensive multi-level evolutionary framework provides an optic that is epistemologically encompassing while remaining teleologically neutral and foundationally uncommitted. It is an optic in which theoretical rigor is not obtained through the imposition of any single mono-causal theory, but through a clear account of how the coevolutionary interpenetration of *multiple* logics actually plays itself out over extended periods of time. Although each of the logics appears to underwrite a distinctive line with quasi-deterministic tendencies, this does not ultimately lead to any general deterministic outcome, since the logics are obliged to perpetually interact.

It is for this reason that I am proposing my nesting dolls, with their quest for a multi-level theoretical synthesis that does not privilege any one dimension of the human experience, but embraces them all by apprehending them additively, through the cumulative lenses of their

---

[6] I discuss the Marxist influences on my own thinking in Koditschek (1996), and Koditschek (2013). The influences from other theories have been less explicit in my published writings. They are simply too numerous to mention here. Suffice it to say that among scholars at work during my lifetime, the works of Anthony Giddens, Michael Mann, W.G. Runciman, Talcott Parsons, Neil Smelser, Charles Tilly, W.E. Sewell, E.A. Wrigley, Peter Mathias, Thomas Kuhn, and Michel Foucault (to cite only several of the most important) have been very influential in my thinking about how other paradigms such as 'Modernization' or 'Geo-political competition' can and could work in the historical sciences.

successively emergent historical origin points. Just as the advent of culture introduced a new logic to the course of human development, displacing the old dynamics of natural selection from center stage - and just as the advent of the state brought a newer logic that displaced the centrality of culture, so the advent of capitalism has brought the newest logic to displace the centrality of the state. Yet, amidst all these dramatic transformations and displacements, neither the state, nor culture, nor natural selection have ever disappeared. Even today, when global capitalism has reached its current pitch of intensity, the nation state remains essential for its operation, and even the most powerful multi-national capitalists must attend to the legal regulations and proscriptions of at least a few powerful states. Correspondingly, both states and capitalists must attend to the residual impact of culture, which can authorize resistances to the exercise of economic or political authority, while also providing channels through which political loyalty can be tapped and economic profits can be made.

Finally, natural selection remains silently at work in the background, though not in the manner envisioned by the nineteenth century racists, sexists, and social Darwinists. On the contrary, it has left us with us with a single common ancestor, whom some have named (half-fancifully) ‘Mitochondrial Eve’. She then begat descendants of enormous (yet not infinite) behavioral plasticity (nearly eight billion alive today), whose minds and bodies are now (in all important respects) of a common design. Natural selection *has* thrown occasional curve-balls at this long line of human creatures in the guise of degraded environments and dangerous epidemics. Moreover, it has specified the institutional and motivational work-arounds that these creatures have needed to invent in order to adapt to the changing cultural, political and social arrangements with which they have successively built their collective habitats.

These behavioral work-arounds and biological feedbacks that humans have crafted to weave diverse local cultures around bodies and minds that change much more slowly are being investigated with considerable energy and imagination by the evolutionary psychologists and anthropologists who study gene/culture coevolution. What we need now is scholars with training in history, historical sociology, political economy, and economic development to join in the evolutionary fray. Even the evolutionary anthropologists who would subsume all human history within the category of ‘culture’ sometimes recognize that this category is too narrow to encompass much that people have built, thought, and done over the past six thousand years. Relations of ‘power’ they reluctantly acknowledge are inadequately treated by models that assume that culture is transmitted over many generations by some combination of vertical and horizontal copying of past practice, modified to a degree by innovation and personal choice (Durham, 1991). For the historical record shows that most ordinary people, since the epoch of the Neolithic revolution, have exercised very limited choice in such matters. They have been told by political (and sometimes economic) bosses the limits of the culture they are permitted to access, and what it will (and will not) allow them to do.

Moreover, 'power', as any historian, economist, or sociologist knows, is not a simple or unitary phenomenon. There is a substantial difference between patriarchal, political, and economic power, while the exercise of power in any of these forms always entails a varying mixture of coercion and consent. Yet, a common feature of all these forms of power is the fact that, over the long run, their scope and shape has been determined by the vicissitudes of state power, which provides essential mechanisms for their enforcement and legitimation. Historians, sociologists and students of law have produced libraries of monographs to illustrate these processes, but they have generally eschewed casting their accounts in a comparative evolutionary frame. At the other end of the academic spectrum, students of modern culture are usually blind to its evolutionary origins. All too often, those who write about contemporary culture ignore its deep roots in the early formative pre-historical development of our species, when many of our most ingrained cultural propensities took root. (For example, consider the essays in anthologies such as Grossberg et. al., ed., 1992 or During, ed., 1993).

The purpose of this paper then is to help devise a trans-disciplinary language in which historians, historical sociologists, and evolutionary anthropologists can start communicating with one another about problems in which we are all mutually interested, but which we understand and describe in very differing ways. If I had to pick one evolutionary term/concept to recommend for active use by social science and historical practitioners, it would be 'adaptation'.[7] Unlike many theoretical concepts which are understandably dismissed as too

---

[7] I am sensitive to the temptation that I should distinguish adaptation as a feature of the evolutionary process from 'adaptationism', as an ideology that assumes almost any and every feature of an organism, a culture, a polity, or a capitalist enterprise can and should be understood as an adaptation to some selective pressure or competitive imperative. It is perfectly possible to acknowledge that many features of the unit in question (organism, culture, polity, or capitalist enterprise) may have arisen accidentally or adventitiously, while still focusing primary attention on those that are adaptive in some way. For analyses that emphasize (often to the point of over-emphasizing) the importance of adaptation as a central concept in biological evolution see Williams (1966) and Dawkins (1986). These views have been persuasively criticized by Gould and Lewontin (1979). Nevertheless, I

broad or too narrow for guidance in empirical study, or too closely tied to a particular mono-causal theory, 'adaptation' is unexceptionable on all these counts. From an evolutionary standpoint, any unit must adapt to the environment in which it functions. This is true, whether the unit is a gene, an organism, a species, a culture, a state, or a capitalist enterprise. At the level of natural selection, adaptation is usually understood as a simple adaptation to nature although, as Richard Lewontin and others have pointed out, many organisms substantially create the niche, which they then occupy (Lewontin, 2000, Odling-Smee, *et al.*, 2003). Humans are preeminently creatures of this sort, and the niches that we first created were our cultures, which have enabled us to successfully occupy every habitat on earth. Note, however, unlike the niche that is constructed by the earthworm, or the beaver, the niches that humans have built are plural, and vary greatly from one habitat to the next. Some bring us to cold places, and others to hot. Some give us access to lush forests, others challenge us to adapt to arid tundras. Some have left us subsisting in small scale communities, while others have driven us into larger units through markets, other distant exchange networks, or the compulsion of central states. Because these variegated social environments favor divergent styles of normative behavior, they have pressed on our innate behavioral plasticity in a host of disparate ways. Some of these 'extended phenotypes' (to use Richard Dawkins's felicitous phrase) draw more heavily on our innate propensity towards cooperation, altruism, and egalitarianism, enjoining us to put a lid on our impulses towards selfishness, competition and hierarchy (Dawkins, 2016). When we find ourselves in other cultures and/or other units of social organization, these behavioral priorities become reversed.

When we look closely from a coevolutionary perspective, we can see that adaptation is no longer a one-way street: It now takes the form of mutual co-adaptation. Individual humans

---

believe that in his campaign against an 'adaptationist program' Gould was sometimes moved to create a straw man, since I am unable to identify any general formula for identifying when the legitimate identification of innovative features as 'adaptations' has gone so far as to become an 'ideology'. My own approach has been broadly shaped by an early and influential reading of Merton (1964) who was concerned a half-century ago to assert the general relevance of concepts of structure and function to all social science, at a time when the term 'structural-functionalism' had been monopolized by one particular school, while those social scientists who were beginning to take the 'interpretive turn' were becoming averse to functionalist explanations of any kind (e.g. Geertz, 1973). In my view, the debate over adaptationism is largely a reprise of the same exaggerated polar oppositions, now transposed from social science into biology.

must adapt to the demands of their cultures, but these cultures must simultaneously serve the collective needs of individual human beings. In the short run it is our ultimately 'fixed' biological natures that seems to set the limits on human adaptation, since cultures will survive only insofar as they serve the concrete needs of real-life humans, whose capabilities for behavioral and bodily plasticity cannot be pushed beyond a certain point. Yet, over the long-run, the logic of culture can become a runaway process, inasmuch as it forces the human mind and body along a path of gradual change. Judging by the behavior of chimpanzees and gorillas, our earliest hominid ancestors were probably far more intrinsically selfish, hierarchical, and less intelligent creatures than most of us are today. But, when atypically high levels of altruism, egalitarianism, cooperation and intelligence turned out to be advantageous to our ancestors in the competition with other early hominid groups, cultural selection favored cooperative, altruistic dispositions in individuals to the point in where random but periodic genetic mutations baked them into the DNA of some future descendants, who gradually spread these (now innate) qualities throughout the gene pool of the species as a whole (Hrdy, 2009; Bowles and Gintis, 2011. See also Richards, 1987, p. 451-503 for a discussion of the related 'Baldwin Effect').

11.5. Analyzing the Logics Separately and Integrating them into an Evolutionary Whole

What happened then, when fully encultured members of species *Homo sapiens* began to add the two additional levels (domination and capitalism) on top of their now entrenched local cultural repertoires? The coevolutionary picture certainly became much more complicated, but the underlying principles of co-adaptation remained the same. Once effectively centralized states made their appearance, the local cultures within their territory (and frontier regions, to a lesser degree) had to adapt to state imperatives, exactions, and commands. As revenue was hived off in taxation, soldiers were conscripted, and local spiritual practices were corralled into state religions, these local cultures felt the pressure of having to acquiesce in submission to rulers with their instruments of central state control. With the advent of large multi-national empires from the Axial Age onward, equally vast and ecumenical universal religions such as Christianity, Islam, Hinduism, and Buddhism sprang up (Eisenstadt, ed., 1983). These universalistic religions shepherded far-flung local cultures into some kind of spiritual coordination such as would be compatible with the imperial civilizations that now encompassed them. Yet, because these diverse cultures had deep roots in their adherents' indigenous ways of life, there were limits to the plasticity that the central state and universalistic religious authorities could demand of them. As a result, these big religions and seemingly all-powerful state bureaucracies had to make repeated accommodations with the local populations that they subsumed. State and imperial culture invariably became a kind of negotiation, in which the center exacted the degree of loyalty and conformity it deemed essential, while leaving local practice to its own devices in many domains (Koditschek, 2019, Homans, 1941, Bede, 1968).

It should be noted that we are now talking about an evolutionary dynamic between states (with their official religions) and cultures (with their ingrained local practices and traditions) in which biology appears to play a negligible role. Yet biology did still set some limits every time hunter-foragers or pastoralists, with their free and egalitarian traditions were called upon to acquiesce in the new levels of higher-order cooperation, submission, and hierarchy that the state (with its class structures and organized religion) demanded of them. It is therefore not surprising that the long history of traditional state-based civilizations is the history of forcible conquest. It is correspondingly, the history of only grudging conditional consent, periodically punctuated by slave revolts, peasant uprisings, savage exterminations, and barbarian eruptions, as well as civil wars, and dynastic overthrows, in which no large imperial state has remained fully intact for more than a few hundred years (Diamond, 1997, Golden, 2011). No sane person will confuse this evolution with 'progress'.

What then happened when the logic (and level) of capitalism was further added to this co-adaptational mix? In the beginning, of course, capitalist enterprises had to accommodate themselves to the imperatives and demands of the states that protected them, allowed them to operate and, increasingly, encouraged them to embark on mercantile and entrepreneurial initiatives on a world-wide scale. However, once states and their consumer populations became sufficiently dependent on these production and trading initiatives, it was capital that began to call the shots. The character and organization of states then had to change to reflect the interests of property-owners and the new incentives of profit maximation. States which failed to introduce these changes became vulnerable to internal revolution, inability to compete militarily, or they became targets for imperial intervention by superior powers who turned them into formal or informal colonies. At the same time, capital continued (and now still continues) to require the power of sympathetic states – to protect capitalist property, to set rules for legitimate competition, to put down insurrections and countervailing collective organization by propertyless classes, and to assist private capital in penetrating those groups, places, or domains that would resist its transformative intrusions. If this is called ‘progress’ then it must be recognized as barbarism to an equal or greater degree (Anderson, 1978, Wolf, 1982, Abernethy, 2000).

In standard accounts of the relationship between capitalism and the state, the role of culture is largely neglected, while the individual human is generally treated as an abstraction – *homo economicus*: an entirely rational, self-interested, individualistic utility maximizer, who acts to optimize his/her resources in a competitive world. Anyone who approaches this situation from an evolutionary perspective will not begin with such dogmatic assumptions. While it is certainly true that capitalist culture favors the competitive, individualistic and self-interested impulses, thereby incentivizing participants to downplay their altruistic, cooperative collectivist sentiments (or at least to relegate them to a subordinate, domestic sphere).[8] Yet, these new values tend to run in opposition to those collectivist, compliant, hierarchical values that were favored by deeply rooted traditional cultures, particularly those that have evolved as a result of a long and successful co-adaptation to the needs of traditional dominance states (Tawney, 1926, Polanyi, 1944, Ostrom, 2015).

When faced with the demand to shift their inner norms to the pro-capitalist valence, such traditional cultural communities have often been inclined to resist or revolt, especially when

---

[8] The role of culture in generating the technology and economic behavior necessary for capitalism has been emphasized (perhaps overemphasized) by Mokyr (1990) and Mokyr (2009). My point is to extend this further to consider the impact of the culture of possessive individualism on reshaping the state and the individual personality as well. I hope this will now clarify the general point that I made in footnotes 3 and 4: In a capitalist society, the pressures to adapt to the logic of capitalism will generally take precedence over pre-existing pressures to adapt to the logic of political domination, or cultural belonging, except where capitalist pressures are so extreme that they would conflict with and jeopardize the fundamental integrity of the polity, the culture, or the biological and psychological imperatives of human nature. In other words, the state, and culture have been (and are likely to continue to be) reshaped more fundamentally by capitalism than is the case the other way around. The more powerful, and more globally extended capitalism becomes, the more this is likely to be the case. Nevertheless, capitalism can never become a totalizing framework for human social life. So long as we remain the creatures that gene/culture coevolution has made us, we will continue to require culture (including such relationships as kinship, community and family) to mediate our ability to survive in a market-driven world. Nor do the current generations of humans show much capacity for abandoning their nation-state identities. What the future will bring, of course remains unknown. Even a fully-developed evolutionary social science is not likely to give us any definitive prognosis about our future. Nevertheless, it might help us to estimate the probabilities as to which configuration of social arrangements are less, likely to be sustainable, and which are most likely to and optimize the quality of human life.

they have been egged on by residual precapitalist elites, or by state authorities who themselves stand to lose by the change (Hobsbawm, 1959). Indeed, the possibility of transforming the social and economic policies of the state often depended on the successful interposition of a political revolution, which only succeeded when it was preceded by a spontaneous cultural revolution in at least a few strategically placed local communities. By various expedients, the inhabitants of these communities began to valorize the new egalitarian, possessive individualist norms for unusual cultural and psychological reasons, well before the regime of capital accumulation was able to promise any large reward in material terms. We need not subscribe to all the details of Max Weber's argument in *The Protestant Ethic and the Spirit of Capitalism*, or Michael Walzer's *Revolution of the Saints*, to see that this role was played by the Calvinists of the Netherlands and by the Puritans in Britain and North America, who did so much to create a distinctively individualistic, competitive culture in which the values and virtues of capitalism could prematurely gestate (Weber, 1958, Walzer, 1965). In France, a similar role seems to have been played by the enlightenment secularists in laying the groundwork for the Revolution of 1789 (Gay, 1966, 1969, Furet,1989, Sewell, 2021).

However, once the values of capitalism have conquered the state, resistant local cultures have come under great pressure to get with the program. In those places, where capitalism and the state have been introduced by culturally alien imperial forces, this traditional cultural resistance has often been formidable. The results have left a trail of blood and tears that has disfigured inter-cultural encounters between indigenes and colonizers, first in North America, then throughout Asia, and still today in Africa, and especially, in the Middle East (Wolf, 1969, Rodney, 1982, Fairbank, 1986, Stannard, 1992, Khalidi, 2004). In this era, when the 'cultural turn' has so completely come to dominate so many of the social sciences and history, it is ironic that so little attention has been played to the critical role of culture in enhancing and catalyzing the spread of capitalism in some places, while resisting and retarding its progress in others. History shows that it is only when possessive individualist values have spread quite far amongst established and insurgent elites that the state will be transformed in a manner that makes it safe for capitalism. In the west, it was only with the replacement of authoritarian/ aristocratic states by quasi-republican constitutional monarchies – in the Netherlands (1648), Britain (1688), the United States, (1783), and France, (1791) that the cultural and political conditions for fully-fledged capitalist expansion were legally formalized (Palmer, 1959, 1964, Rudé, 1964, Wilson, 1968, Hill, 1969).

Moreover, as we have already seen, enormous pressure has been exerted on local cultures from the very first appearance of superordinate political entities. The advent of state religions, culminating in the great ecumenical faiths of the Axial Age, can be viewed from a coevolutionary angle as the construction of a set of structures and institutions (i.e. priesthoods, theological texts, oral catechisms, and clerical hierarchies) that were designed to shape the adaptation of indigenous local cultures to the enlarged political forms. Indeed, these churches and faiths were the primary adaptive instruments that were designed to reconcile individual subjects to consensual acquiescence in authoritarian political regimes. At the same time, these churches and faiths served as more or less effective intermediaries between rulers and subjects. Again and again, they found a distinctive role in trying to convince subjects that it was in their spiritual interest to obey the rulers, while convincing rulers that it was in their practical interest not to press the demands of exploitation too harshly, and to accept the inevitability of a good deal of local plebeian autonomy (Homans, 1941, Kantorowicz, 1957, Le Roy Ladurie, 1978, Eisenstadt, ed., 1983).

Although the requirements of religion changed with the advent of capitalism, it has continued to play an important cultural role, sometimes reinforcing the behavioral imperatives of possessive individualism (as with Calvinism) or by seeking to temper the drive to productive accumulation by emphasizing the necessity of a separate (albeit usually domestic) reproductive, restorative sphere (Sklar, 1976, Douglas, 1977, Ryan 1981, Davidoff and Hall, 1987). As capitalism has pressed on to its more mature, abundant phases, capitalist culture has tended to grow increasingly secularized, shifting its focus from the imperatives of production to the enticements and allurements of consumerism. Where political centralization introduced the need for an initial 'scaling up' of culture, consumer capitalism has scaled it up in a different way. As brand-name goods have become globally iconic, the influence of ubiquitous advertising has created imagined communities of aspirational shoppers, which sometimes extend around the world. Here of course, the advent of twentieth century mass media and mass communications – first newspapers and magazines, then film, radio, television, and the internet – have laid unprecedented new technological foundations for an almost limitless augmentation of this cultural 'scaling up' (Postman, 1985, Lears, 1994, Cohen, 2003).

Indeed, we now inhabit a world in which culture is no longer primarily local, since it is constituted by behavioral norms, fashion trends, and consumerist affinities, whose substantive content can change from week to week, and whose reach has extended across every continent.

In that sense, this hyper-modern capitalist consumer culture seems antithetical to the traditional forms of culture which drew their strength from being time-honored, venerated, and local. The boosters of this vast homogenizing culture vaunt its liberating potential, as opening the way for new, individualized identities. Yet when the culturally liberated individual feels exploited or lost in anomie, s(he) becomes fodder for authoritarian fantasies. Here too, the slippage between an aspiration towards freedom and a descent into reaction masks many co-adaptive compromises: Old, entrenched, gendered norms clothe themselves in seductive new imagery. Newly fabricated 'invented traditions' conjure up nostalgic longing for the verities of yore (Allen, 1984, Faludi, 1991, Hochschild, 2016).

11.6. Integrating Variational with Transformational Processes: Dynamic Equilibrium, and the Relationship of Continuity to Change

These reflections on the scaling-up of culture warrant some further observations about a more general pattern of 'scaling-up' that is characteristic of the recent phases of human coevolution, particularly those activated by the logic of domination and the logic of capitalism. For these most recently emergent levels are consequentially different from the earlier levels that were set in motion by the advent of culture, and the operation of natural selection on the minds and bodies of early human beings. Individual persons and (slightly more problematically) the constituents of simple, local primordial human cultures, can be represented as discrete units of coevolution. In that sense, they fit reasonably well into an evolutionary theory that is, like the theory of Darwin, dependent on a sequence of variation, selection, and inheritance to pass on characteristics (in altered distribution) from one generation to the next. Yet, in a series of trenchant critiques, the eminent population geneticist Richard Lewontin questioned whether gene/culture coevolution really operated this way. All meaningful social evolutionary theories, he contended, have in fact been theories of a very different type. So far from being based on discrete units susceptible to variation, selection, and inheritance, they were internalist studies of some large, composite but singular societal entity that was undergoing dynamic developmental change. As a result, they aimed to trace *transformational* changes *within* the life-course of this entity, rather than the *variational* permutations *between* units that are competing with one another (Fracchia and Lewontin, 1999, Lewontin, 2012).

Lewontin makes an important distinction between variational and transformational approaches to evolution. Yet, he errs in assuming that we must choose between the two as incompatible alternatives. One of the advantages of my multi-level approach to coevolution is that it now becomes possible to shuttle between these alternate perspectives when examining any two levels together. Consider the coadaptation between local cultures and central states that was discussed in the last section. Here we have entities at two levels, each of which is being internally transformed by variational selections that are also operating on both levels: Local cultures must adapt to a transforming state (that is in competition with other states, and with the desire of constituent cultures to retain autonomy). To achieve an evolutionarily stable outcome it is necessary for units at the lower level (local cultures) to vary in a manner such that the more state-accommodating variants can be selected as adapted to the needs of a self-transforming state which wants/needs to enhance its internal coherence. Yet, at the same time, those variants of state forms, which develop the most effective state churches are likely to be favored in selective competition with other states because of their ability to incorporate those local cultures that have been transformed to accept (without entirely capitulating to) state authority.

The analogy with nested dolls may be helpful in restating this in the language of general principle. At this stage of the analysis, two of the levels (dolls) have been separated, so that some particular feature in which both are implicated can be dissected, and the two relevant interacting variables can be extracted. Which cultural variants are most conducive to strengthening the state? Cultural variants are now the independent variable, and strengthening the state is dependent on selecting the right one. But then the analysis is reversed. Which state forms or polies can most readily adapt to cultural invariants (dependent variables, now held constant). When these reciprocal operations are completed, the dolls can be put back together in their nested relationship for a fresh set of analytical questions to be asked.

From one point of view, it may seem like an artificial exercise to isolate the specific multi-level relationship between variation and transformation at a single frozen moment in time. Our ultimate goal, after all, is to understand their dynamic inter-relationship as they proceed in tandem through the evolutionary flux. For historians this is clearly connected with the much more familiar problem of continuity vs. change: Which is deemed to be more salient in any given time/place circumstances, and what is the relationship between the two? The evolutionary approach offers a means of reframing these questions in a more analytically precise and empirically sensitive way. At any given moment, one factor is shown to be changing relative to continuity in the other. In other words, the evolutionary approach is grounded in the premise that elements of continuity (reproduction) are always present in the flux of change (transformation), simply because they involve variations that are being selected (i.e. mutant adaptations) in which existing structures have been retooled (i.e. transformed) to perform new functions.

In biological evolution, such episodes of phylogenetic change are analyzed as one-step processes that ultimately lead to an evolutionarily stable strategy. Organisms become maladapted as environments change. Organism and environment fall into disequilibrium, but the equilibrium is restored through adaptation (Maynard Smith, 1975). However, in the case of humans, where multiple evolutionary levels are operating together, there can be no enduring re-stabilization, because re-stabilization at one level is likely to precipitate destabilization at another. This is the case for three reasons: 1) As the logics of culture, domination, and capitalism become increasingly operative, the environment to which humans must adapt becomes a social environment – i.e. one in which nature is no longer purely external to the human experience, since it now incorporates a 'second nature' consisting of social practices and institutions that activities of previous generations have brought into being. Consequently 2) The evolutionary trajectories laid out by the logics of culture, domination and capitalism ensure that every equilibrium will be a dynamic equilibrium, since the goalposts that specify the demands of adaptation will constantly be moving. 3) Since the timescales on which the evolutionary dynamic is moving generally speeds up through each of the three added levels, most equilibria achieved at the level of culture will eventually be destabilized by more rapid change at the level of the polity that is causing disequilibration at the levels below. Eventually this disparity in the pacing of evolutionary change is further exacerbated by the logic of capitalism, which is introducing disruptive factors even more rapidly than polities or cultures can generate adaptations to the new conditions.

The result is that the human world is always changing. But this does not mean that threads of continuity are merely vestigial. It simply means that the dynamics of evolution are always moving too rapidly for adaptations to solidify and to achieve an enduring evolutionarily stable state. Since destabilizing factors are always jeopardizing any evolutionary equilibrium, there will always be new challenges, which will require the social evolutionist to perform another analysis of another moment in the rebalancing between continuity and change.

11.7. Some Examples of How this has Worked: Coming to Terms with the Problem of Agency

In our analysis, we gave the example of the rise of Calvinist culture in certain sixteenth century European localities. From the perspective of the traditional patrimonial state, this development was certainly dangerous and maladaptive. Yet, Calvinism survived, and its culture spread because of the way in which it proved co-adaptive to the needs of nascent capitalism, which was just then poised (with the help of Calvinist culture) to begin transforming the traditional patrimonial state (Walzer, 1965, Wilson, 2002). In this case, our scenario has grown more complicated with the addition of a third level of explanatory logic (the level of capitalist developmental logic) that has been added to the pre-existing two. What happens when we keep in mind all the background interactions of all the relevant operant logics including, in this case, the logic of genetic and individual personality selection/transformation)? Indeed, while there is no reason to think that genes played any significant role in this instance, it is the clear that the success of Calvinism (and, by extension, of embryonic capitalism and the early capitalist state) did depend on the spread of a new personality type of 'worldly asceticism' that Weber (and, on a political axis, Walzer) described (Weber, 1958, Walzer, 1965). So, it should be clear in almost any other multi-level scenario that we should similarly consider how the transformation of any given culture will affect the selection of variable individual personality traits that are conducive to success within it, just as the transformation of any given polity will make certain cultural variants more or less satisfactory in replicating its constituents and their behaviors from one generation to the next. It is hardly an accident that the advent of capitalism has brought personal character traits of possessive individualism to every other level (state and culture) in a manner that has tended to crowd out more cooperative, altruistic personality characteristics. Carried too far, this trend has repeatedly threatened to undermine the fabric of capitalist society. And yet, it is a pattern that the logic of capitalism cannot do without.

It is very likely that the further we pursue this analytical shuttling between variational and transformational perspectives in understanding the dynamics of multi-level social evolution, the more we will be reminded that such a fully elaborated dynamic looks a good deal more complicated than the simple Darwinian model of natural selection that characterizes the study of evolution in the biological realm. In biology, the internal processes of ontogenetic organismal development have conventionally been treated as completely distinct from the variational dynamics of phylogenetic species evolution. Nevertheless, as Ariew and Panchanathan show in this volume, even in the domain of pure biology, this absolute distinction becomes a bit hazy once one begins to view organisms not merely as mechanisms, but as agents who can affect their own development, thereby subtly altering the long-term course of their species evolution as well. (In situations like this both ontogenetic development and phylogenetic evolution are implicated in the concept of 'niche construction').

This complexity becomes even more consequential after we move into the domain of human evolution, where the agency of individuals in collectively shaping and reshaping cultures is far more impactful than with any other species. In contrast to other organisms, human agents act consciously (and usually in a purposive manner) in selecting the cultural variants that they favor. When we add the additional levels of domination and capitalism to the coevolutionary pathway(s), the effects of human agency become even more consequential and impactful than when biology is considered alone. Once we begin to treat cultures as variational units at the next evolutionary level, the consequences of conscious choice in values and practices becomes transparent to all concerned. We can also see why the pressures for change appear more rapidly as we move up the levels/logics of the social evolutionary chain. By modern standards, cultural change starts out slowly when it has only to respond to external environmental changes (e.g. changes in local ecologies or interactions with other cultural groups). Under these circumstances, cultural selection is most often made blindly, by following prestige bias, or the inertia of custom and tradition (Richerson and Boyd, 2005, Henrich, 2016).

By contrast, the interposition of the political level/logic of the organized state puts pressure on cultural change to speed up. States will succeed or fail in competition with others by adopting innovations that are recognized to be compulsory, especially in the areas of military technology and centralized bureaucratic control. As a result, successful states (unlike simple, informal cultures) quickly evolve formal, hierarchical, structures in which political decision-making powers (i.e. powers of selection) are monopolized in the hands of a sovereign governing elite. Since the whole point of this institutional innovation has been to make states more optimally competitive variational units, the transformations that are introduced internally by policymakers will *not* be random. Rather, they will be geared with maximum efficiency towards the ability of any given political unit to compete with other (purposively self-transforming) states (Koditschek, 2019).

Adding the level of capitalist competition further into this coevolutionary mix, the explanation of any specific multi-level scenario will become complicated in an even more entangled manner. Like state policymakers, capitalist entrepreneurs, managers and investors must make conscious decisions (selections of strategic variants) based on the dictates of market competition. However, because polities are needed to set the conditions for market competition, the rules of the market are partly shaped by state interest, just as the imperatives of market competition increasingly set terms for the coevolution of the state (as well as of the cultures contained within it). What then is the basic unit of capitalist competition (i.e. selection)? To some degree it is the individual competitive firm. But, on another level, it becomes the unique state versions of capitalism that are themselves the products of a complex evolutionary prehistory. As capitalism becomes ever more global, these state capitalist variants compete with one another for global hegemony, even as they try to carve out quasi-autonomous zones in which they set out rules of the market competitive game. This was the situation during the 16th to 18th century era of competing state variants of mercantilism, until the conditionally liberal British imperial model triumphed in the nineteenth century. It became again the situation during the twentieth century, when the British model was challenged first by a German reactionary-authoritarian model, then by a Soviet communist-authoritarian model, and finally an even more expansive (and presumptively liberal) American imperial regime. After 1945, the American model appeared to win out across the globe, as it became hegemonic and increasingly compulsory. Yet after unleashing a new (postcolonial) wave of fully global capitalist development since the 1970s, American hegemony now finds itself under pressure, from a newly energized Chinese demi-authoritarian model on one side, and an entrenched (but perhaps more ultimately adaptive) European social democratic model on the other (Koditschek, 2019).

11.8. Retooling Theories of Social Transformation for the Social Evolutionary Enterprise: Integrating the Analysis of Structure with the Impact of Agency

Clearly, close attention to the ways in which various logics interact calls for a much deeper enquiry into the nature and force of the logics themselves. Such an enquiry cannot be pursued very far in a paper that has already grown too long. However, the question can safely be deferred in light of my observation in section IV that these logics should be understood as postfoundational modifications of a series of long familiar social theoretical paradigms that have dominated the study of transformation in the (non-evolutionary) social sciences for some time. These paradigm proposals enjoyed considerable prestige during the third quarter of the twentieth century, but they are now generally rejected for their tendencies towards monocausal reductionist explanations, and for their contamination with teleological notions of progress. In particular, when contemporary historians and historical social scientists approach their study of the modern period, they are frequently moved to reject these large transformational theories that were extremely popular five decades ago. This is particularly true with regard to Modernization, which explains the rise of capitalism in a way that generally affirms and legitimizes it, and of Marxism, which offers a transcendent, dialectical critique of the Modernization analysis of capitalism and of its fate. Indeed, it is consciousness of the flaws in these theories that animate much of today's suspicion of evolutionary social science.

Recognizing that the actual pathways of variable state-capitalist development have differed from these universalizing theories in key respects, most conventional historians today retreat into a sui-generis transformational nominalism that would treat each case a unique story unto itself. Yet, as I argued in section IV, by relocating the most valid (and empirically supported) elements of Modernization and Marxist theory into a multi-level evolutionary framework, it should be possible to separate the baby of genuine transformational insight from the bathwater of dogmatic reductionism and arbitrary teleology. When we transpose the particularistic stories told by nominalist historians into the generalizing idiom of multi-level evolutionary analysis, what previously appeared as idiosyncratic anomalies can be reconceived as hybrid trajectories based on compounded conformations. The distinct narrative accounts of different time/place transformations can be re-written as the relevant set of competing variants that is favored or disfavored by the selection process.

Under the logic of domination this will seek to identify the range of competing alternative state systems (and their constituent cultures and individual persons) that are vying for dominance in a geopolitical frame. Under the logic of capitalism, it will look for the ensemble of state, cultural, and individual strategies that are contributing to the development and metamorphosis of the capitalist system as a whole. Among the features it will surely light upon will be the pattern by which capitalism has advanced through successive cycles of boom and bust. As it has drawn ever more cultures, polities, and regions into the snares of its commodity exchange web, markets have expanded, consumers have been enlisted, products have been transformed, and workers have been proletarianized. Workshops were replaced by factories, which were then supplanted by multi-national chains of disaggregated production and, most recently, by electronically linked networks of informational goods.

At any given moment in this rapidly shifting capitalist landscape, different strategies have proven more optimally adaptive, as the initiative has passed from one nation-state-inflected model to the next. In contrast to the claims of the dogmatic transformational theories that posit one true road to the future, the multilevel evolutionary perspective (embracing differential pathways of transformation, but also invoking the dynamics of variation, selection and reproduction) remains open and contingent. It objectively registers the reality that some innovations (variations) have proven beneficial, while others have been modified, abandoned, or have led to the dead end of failure. Each capitalist state variant has correspondingly come to the competition burdened with the ballast of its unique evolutionary past. Accumulated national traditions have repeatedly been sorted, sometimes modified, and sometimes abandoned. In some cases, they have awkwardly endured as impediments to development, while in others they have fortuitously re-emerged as spandrels or exaptations that promote adaptive optimality in unanticipated ways (Moore, 1966, Gould and Lewontin, 1979). In a manner that is eerily similar to biological evolution, historical evolution exhibits a kind of opportunistic bricolage. In this *ad hoc* fashion, structures that have lost their former functions accumulate adaptations that fit them for new ones, while the challenges of a changed environment can be met because the social organism is able to retool vestigial features to meet the altered demands of a different time.[9]

None of this would be possible if social historical evolution were unable to draw (perhaps in a less purely random fashion) on the same expedient accidents that natural selection seizes upon to enhance the adaptation of biological organisms. Were we to restrict ourselves to conventional (highly directed) theories of transformational (internal) evolution (as Lewontin argued, we must) we would always be in danger of getting trapped in one perspective. This

---

[9] Here, I am quite deliberately retrofitting for the study of social evolution arguments originally made by Gould and Lewontin (1979) for biological evolution.

would render us intellectually impoverished in the face of unexpected developments — as it also leaves us prisoner to some value laden notion of 'progressive' teleology. By insisting on combining a transformational with a variational standpoint, however, we can avoid this outcome. To be sure, the strategically mapped developmental trajectories of successful states and capitalist units do impart an element of directionality to social evolution, simply because these have turned out to be the strategies of success. However, variational competition and selection in a complex environment always ensure that no particular (theoretically predetermined) outcome will ever permanently be achieved. Our dialectical approach to social evolution (simultaneously both variational and transformational) entails a recognition that the future is indeterminate. Whatever the goals of any particular actor (or theorist), it is necessary to embrace the possibility that things may go awry. Moreover, because this dialectical perspective encourages detachment from single self-interested viewpoints, it encourages the understanding that what represents progress for one player is likely to constitute disaster for another. This does not mean that the evolutionary scholar must abandon his/her personal ethical and ideological commitments. It does mean that he/she must situate them realistically in the more crowded and constrained coevolutionary ensemble in which any program for the future is not guaranteed by some telos, but must earn its (contingency dependent) way.

Another further advantage of this multi-level approach to evolution is that its emphasis on a succession of increasingly nested social structures also draws out important implications for the role of human agency. At the level of simple autonomous cultures there can be in fact no systemic contradiction between the structures of the cultural order and the exercise of agency. According to the models of the Boyd, Richerson, Henrich school of evolutionary anthropology there is a strong tendency for the customs and traditions of one generation to be passed on (perhaps with gradual modification) to the next, except where stronger environmental pressures (or competition with other groups) induces adaptive cultural change.[10] In these models, cultural

---

[10] Richerson and Boyd (2005) as well as Henrich (2016) emphasize the important consequences of conformist bias and of prestige bias in cultural transmission. As an historian, I find this theoretical hypothesis to be useful in understanding why traditional cultures tend to be generally conservative and worshipful of custom and tradition. However, I think the situation is far more complicated in the modern capitalist era than these evolutionary anthropologists allow. For under conditions of modern capitalism, a different default option becomes widepread, in which strong adaptive pressures motivate frequent and systemic cultural change. Yet, here we find the curious irony that conformist bias and prestige bias are often hijacked by the shapers of public opinion to valorize

selection is assumed to be the province of free individuals. Although such models likely underestimate the impact of misfits, doubters, rebels, and innovators, they do generally conform to what we know about simple foraging cultures, where the operation of subjective agency rarely produces many enduring departures from the adaptive imperatives of structural continuity (or, where necessary) change.

Once the logic of domination begins to operate however, individual agency becomes radically constrained. According to the norms (now laws) of the state, its proper exercise is confined to the interests and desires of the system's male elites. Women, children, servants, slaves, and the unmarried are expected to acquiesce in a status of permanent subordination. Married men of the lower classes are 'bought off' because they are usually permitted some modicum of patriarchal authority as household heads at the local cultural level in return for eschewing all claims to meaningful political power. Under these circumstances, the only way that peasants, workers, women, and other household dependents can exert any overt political muscle is through movements of collective organization and agency. Such collective movements are difficult to mount and even more difficult to sustain. Occasionally they erupt at the local level, where they can be easily crushed by the central power. On the rare occasions when they spread like wildfire through the countryside they can inflict spectacular damage, but eventually burn out. On the very rare occasions when they actually succeed in overthrowing established authority, they end up simply replacing the old elites with substitutes of plebeian origin who end up drifting into the same familiar elitist roles (Hilton, 1973, Fairbank, 1985).

The advent of early capitalism does significantly alter this situation, finally building up the accumulated grievances of multiple constituencies to the point where revolts can be turned into revolutions, which finally (though often circuitously) eventuate in the overthrow of old regimes (Rudé, 1964, Stone, 1972). Yet, the logic of capitalism engenders new movements of popular opposition which can, in time, turn into more effective collective movements of resistance and and revolt. Under the freedoms permitted by a liberal capitalist regime, those who are

---

authorized change and innovation (especially in entrepreneur-worship, consumerism, entertainment, and fashion), while anathematizing radical programs for the reform of capitalism itself (Postman, 1985, Herman and Chomsky, 2002).

disgruntled can organize voluntary associations, trade unions, and political parties that offer alternative collectivist visions of a post-capitalist regime (Thompson, 1963, Tilly 1995). Where such movements have precipitated 'proletarian' revolutions they have generally degenerated into authoritarian dictatorships, and when they have turned reformist, they have been incorporated into the system as countervailing powers which temper capitalist market freedom with a tincture of social democracy. But, rarely confining themselves to the strictly proletarian grievances of wage workers, they have often blossomed into wider popular movements advocating for more generalized social justice, drawing attention to discrimination against women, racial minorities and sexual 'deviants'. At that point, unfortunately, they may be prone to dividing the proletariat in ways that ultimately reinforce the established political authorities of neo-liberal capitalist regimes. Where the popular forces of progressive reform have grown irresistible – and are yet counter-balanced by formidable reactionary elements with residual privileges to defend – the result can be a descent into fascist (or quasi-fascist) regimes, although such solutions rarely provide the requisite stability and flexibility for further capitalist development over the long run (Mayer, 1971). In short, there is no single 'inevitable' outcome when the fine balance of multiply configured (and rapidly changing) forces can be turned on a dime in different ways.

Perhaps this is why the ideologies of both capitalism and communism, which each promised inevitable paths to freedom and democracy, have in different ways resulted in updated modes of coercion and in the replacement of old with new elites. Of course, when we contemplate the future, we must insist that this outcome is also not foreordained. Agents are free to advance their favored programs for improved social reorganization. Yet, they would do well to recognize that no particular configuration of cultural, with political, with social, and economic forms must necessarily provide a royal road to human flourishing, prosperity, and universal equality. As so many times in the past, the multi-level approach to the evolution of agency under capitalism, as under domination before it, requires us to accommodate our explanations to the contradictory nature of these systems without ever being absolutely certain that we have taken all the relevant factors into account.

Yet, while this uncertainty strikes a blow against strict determinism, it also enables us to embrace the vicissitudes of agency so as to recount the history of adaptation as a two-way street. Just as rebels and subordinates must adapt to the power of their social superiors, the ruling classes must also adapt to the demands of the downtrodden and exploited whenever collective organization empowers these people to exercise their agency in a systemically consequential way. At the same time, such movements are often co-opted into the system, while elite failures to respond effectively to their challenges will usually result in the collapse of the established regime. Perhaps someday 'the people' will finally be able to come into their (hunter-gatherer cultural) inheritance and find some form of polity and economy that will them to recreate, under modern, globally interconnected conditions, a form of self-government that dispenses with inequality and dominance by elites.

Our multi-level evolutionary framework can give full play to agency of all actors – both individual and collective -- as they constantly intervene to alter the inherited structures, which simultaneously constrain them to a greater or lesser degree. Even the best-laid designs of the most farseeing statesmen, and the cleverest prognostications of the most innovative capitalist entrepreneurs, or the most deeply committed goals of popular movements for reform will always encounter a series of accidental events – i.e. events that s(he)/they perceive as accidental, since the logic of these events rises to the surface from a different level to whose operation s(he)they had hitherto been blind. Such contingencies will necessarily introduce unknown and unanticipated possibilities that must defy the best calculated predictions and upset the best laid plans (Sewell, 2005).

What this means is that the evolutionary approach to human history, like the evolutionary approach to natural history, cannot expect to lay the foundations for a fully predictive science. For the foreseeable future, we must be satisfied with such insights as can be gained from a retrodictive approach. But retrodiction has always served as the intellectual anteroom to non-reductive strategies for facing a future that must remain unknown. Ever since Thucydides proclaimed that "history is philosophy teaching by examples," this has always been recognized as the path to wisdom for anyone attentive to the vicissitudes of human life. By elevating history into a multi-level evolutionary social science we would simply carry such exercises in intuitive retrodiction forward in more precise, comprehensive, and methodologically sophisticated ways. Yet, because evolutionary social science promises us only imperfect tools for anticipating an indeterminate future, it is a scientific approach that will surely end up looking more like meteorology than like nuclear physics.

Unlike physics, meteorological forecasting cannot predict a precise outcome to the third or fourth decimal place. At best, it can draw on massive data from myriad empirical observations of past events to provide a rough estimate of the general direction that is likely to be taken in the future, provided some completely unanticipated event does not throw the calculations off. Weather forecasting today has become a highly data driven and sophisticated scientific activity, drawing on long experience in understanding how atmospheric and hydrological processes work. Billions of discrete observations of events near and distant in time and space are brought together by massively power-driven computer models that process the data, in conjunction with the theory that has been devised to guide the computations. There is no reason, in principle, why evolutionary social science could not reach a comparable degree of sophistication if sufficient resources were devoted to pursuing it, and sufficient wisdom were accumulated by decades of interaction between evidence and theory. However, we are not there yet, and we have a very, very long way to go.[11]

11.9. Conclusion

Kevin Laland and his fellow gene-culture coevolutionists have made a valuable contribution to our understanding of how human social evolution works. Yet, when he invokes the electric pylons, cars, brick buildings and chimneys outside his window, he can do no more than gesture towards explaining a set of phenomena that he and his colleagues lack the tools to understand. My paper makes no pretense of being able to fill in the requisite details in a manner that would turn Laland's industrial (or is it postindustrial?) scene into something more intellectually akin to Darwin's tangled bank. However, I have tried to specify some of the features that will need to be added to the gene/culture coevolutionary framework so that it will be able to take account

---

[11] This raises the question as to whether the multi-level evolutionary ensemble that I am here describing might be analyzed through the formal mathematical models and/or agent-based computer simulations, such as meteorologists and other quantitatively trained scientists have successfully applied to that class of integrative ensembles known as 'complex adaptive systems'. Such systems, though fully analyzable, become ultimately unpredictable, since small perturbations in initial conditions lead to large (chaotic) indeterminacies in long-term outcomes (Glieck, 1987, Gell-Mann, 1994). I am insufficiently conversant in these techniques to offer any judgement. However, my reading of Miller and Page (2007) suggests that the social evolutionary process herein described is indeed a complex adaptive system of this type. I strongly suspect that there will be a good deal of 'low hanging fruit' here for some future appropriately trained scholar, who combines a deep qualitative knowledge of the empirical details of social-historical processes with proficiency in the requisite mathematical and computer simulation techniques. The mission of the Santa Fé Institute seems to be well designed for such an endeavor as laid out in in Pines, (2019). An initial stab at extending this mission to the analysis of human history can be found in (Krakauer, Gaddis, and Pomeranz, 2017).

of such fundamental new emergent properties of social organization as that of state domination and exponential capitalist growth. Where Laland can only point to the pylons, cars and chimneys, such an expanded evolutionary framework would seek to explain how they actually came into existence, and why they were situated in some places rather than in others, and who controls them, to what end? It will need to show why they sometimes appear in the guise of advanced technology -- clean, green, and energy efficient -- whereas elsewhere they are old and rusted -- exuding an odor of deindustrialized neglect. What do we make of those pylons, cars, and chimneys, when we find them available only in short supply, connected to the homes and businesses of the powerful and rich? Should we anticipate that this fragile infrastructure will soon be invaded by squatters? Will it be purloined by pirates, blown up by terrorists, or expropriated by revolutionaries? Or will the poor, the hungry, the unemployed and the overworked be left to tangle with one another on the banks of some polluted stream?

Evolutionary theory will probably never be able to answer these questions definitively, any more than Darwin's vision could tell him exactly how his garden landscape would look in a million years. But this need not be our goal. For we are integrally a part of our social landscape, whereas Darwin semi-detached himself from the one he so eloquently described. His techniques of artificial selection were undertaken as secondary tools to give him intellectual purchase into the process of natural selection, to which he related himself in only the most abstract sense. For us, the active builders of our habitat and constructors of our niche, there is nothing artificial about the selections we make.For it is largely within our power (perhaps not individually, but collectively) to decide on the coevolutionary path that our species will ultimately follow. Will we harness our nature to redirect our purpose, as we have already done several times in the past – re-connecting with that broader nature from which our most recent phases of evolution have left us estranged? Will we re-calibrate these self-made habitats to the nature we have inherited, or will the hubris we have won through our transformations of nature bequeath us only a habitat that drives us extinct?

References

Abernethy, David B., The Dynamics of Global Dominance: European Overseas Empires, (Yale University Press, 2000)

Allen, William S. The Nazi Seizure of Power: The Experience of a Single German Town: 1922-1945, ((Franklin Watts, New York, 1984)

Anderson, Perry, Passages from Antiquity into Feudalism, (Verso, London, 1978)

Anderson, Perry, Lineages of the Absolutist State, (Verso, London, 1979)

Aston Trevor, ed., Crisis in Europe: 1560-1660, (Doubleday, Garden City, New York, 1965

Baines, Edward, History of the Cotton Manufacture in Great Britain, (Fisher and Jackson, London, 1835)

Bede, A History of the English Church and People, (Penguin, Harmondsworth, [731], 1968)

Blaut, J.M., The Colonizer's Model of the World: Geographical Diffusion and Eurocentric History, (Guilford Press, New York, 1993)

Bowler, Peter, The Eclipse of Darwinism, (Johns Hopkins University Press, Baltimore, 1983)

Bowles, Samuel, and Gintis, Herbert, A Cooperative Species: Human Reciprocity and its Evolution, (Princeton University Press, Princeton, 2011)

Braudel, Fernand, The Mediterranean and the Mediterranean World in the Age of Philip II, 2 vols., (Harper, 1966)

Brown, Cynthia Stokes, Big History: From the Big Bang to the Present, (New Press, New York, 2007)

Cavalli-Sforza, L.L., and Feldman, M.W., Cultural Transmission and Evolution, (Princeton University Press, Princeton, 1981)

Chapais, Bernard, Primeval Kinship: How Pair Bonding Gave Birth to Human Society, (Harvard University Press, Cambridge Massachusetts, 2008)

Christian, David, Maps of Time: An Introduction to Big History, (University of California Press, Berkeley, 2004)

Cline, Eric H., 1777 B.C.: The Year Civilization Collapsed, (Princeton University Press, Princeton, 2014)

Cohen, Lizabeth, A Consumers' Republic: The Politics of Mass Consumption in Postwar America, (Vintage, New York, 2003)

Crosby, Alfred W. Ecological Imperialism: The Biological Expansion of Europe, (Cambridge University Press, 1986

Crosby, Alfred W. The Columbian Exchange: Biological and Cultural Consequences of 1492, (Greenwood Press, Westport Connecticut, 1972)

Crosby, Alfred W., Ecological Imperialism: The Biological Expansion of Europe, (Cambridge University Press, Cambridge, 1986)

Darwin, Charles, Darwin, Texts, Commentary, ed., Phillip Appleman, (Norton, New York, 2001)

Davidoff, Leonore, and Hall, Catherine, Family Fortunes: Men and Women of the English Middle Classes: 1780-1850, (University of Chicago Press, Chicago, 1987)

Dawkins, Richard, The Blind Watchmaker, (Oxford University Press, 1986)

Dawkins, Richard, The Extended Phenotype, (Oxford University Press, Oxford, 2016)

De Vries, Jan, The Economy of Europe in an Age of Crisis: 1600-1750, (Cambridge University Press, Cambridge, 1976)

Deane, Phyllis, and Cole, W.A., British Economic Growth: 1688-1959, (Cambridge University Press, Cambridge, 1969)

Degler, Carl, In Search of Human Nature: The Decline and Revival of Darwinism in American Social Thought, (Oxford University Press, Oxford, 1991)

Diamond, Jared, Guns Germs and Steel: The Fates of Human Societies, (Norton, New York, 1997)

Douglas, Ann, The Feminization of American Culture, (Knopf, New York, 1977)

Durham, William, Coevolution: Genes, Culture and Human Diversity, (Stanford University Press, Stanford, 1991)

During, Simon, ed., The Cultural Studies Reader, (Routledge, London, 1993)

Eckberg, Douglas Lee and Hill, Lester, “The Paradigm Concept and Sociology: A Critical Review,” American Sociological Review, (14/6, Dec. 1979), p 925-37.

Eisenstadt, S.N., ed., The Origins and Diversity of Axial Age Civilizations, (State University of New York Press, Albany, 1983)

Fairbank, John, The Great Chinese Revolution: 1800 – 1985, (Harper, New York, 1986)

Faludi, Susan, Backlash: The Undeclared War Against American Women, (Crown, New York, 1991)

Findlay, Ronald, and O'Rourke, Kevin H., Power and Plenty: Trade, War, and the World Economy in the Second Millennium, (Princeton University Press, 2007)

Fiske, Alan Page, Structures of Social Life: The Four Elementary Forms of Human Relations, (Free Press, New York, 1991)

Foucault, Michel, Discipline and Punish: The Birth of the Prison, (Vintage, New York, 1979)

Fox-Genovese, Elizabeth, Genovese, Eugene, Fruits of Merchant Capital: Slavery and Bourgeois Property in the Rise and Expansion of Capitalism, (Oxford University Press, 1983)

Fracchia, Joseph, and Lewontin, Richard, "Does Culture Evolve?" (History and Theory, 38/4, December, 1999), p. 52-78

Furet, François, Interpreting the French Revolution, (Cambridge University Press, Cambridge, 1981)

Gay, Peter, The Enlightenment, 2 vols., (Random House, New York, 1966, Norton, New York, 1969)

Geertz, Clifford, The Interpretation of Cultures, (Basic Books, New York, 1973)

Gell-Mann, Murray, The Quark and the Jaguar: Adventures in the Simple and the Complex, (W.H. Freeman, New York, 1994)

Gleick, James, Chaos: Making a New Science, (Penguin Books, New York, 1987)

Golden, Peter B., Central Asia in World History, (Cambridge University Press, Cambridge, 2011)

Gordon, Robert J., The Rise and Fall of American Growth, (Princeton University Press, Princeton, 2016)

Gould, S.J., and Lewontin, Richard, "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Program," Proceedings of the Royal Society of London, B, 205, (1979), p. 581-598.

Graeber, David, Debt: The First 5,000 Years, (Melville House, New York, 2011).

Grossberg, Lawrence, Nelson, Gary, Treichler, Paula, eds., Cultural Studies, (Routledge, New York, 1992)

Harari, Yuval Noah, Sapiens: A Brief History of Humankind, (Harper, New York, 2015)

Henrich, Joseph, The Secret of our Success: How Culture is Driving Human Evolution, Domesticating our Species, and Making us Smarter, (Princeton University Press, Princeton, 2016)

Herman, Edward S., and Chomsky, Noam, Manufacturing Consent: The Political Economy of Mass Media, (Pantheon, New York, 2002)

Hill, Christopher, From Reformation to industrial Revolution, (Penguin, Harmondsworth, Middlesex, 1969)

Hilton, Rodney, Bond Men Made Free,: Medieval Peasant Movements and the English Rising of 1381, Methuen, London, 1973)

Hilton, Rodney ed., The Transition from Feudalism to Capitalism, (Verso, London, 1978)

Hobsbawm, E.J., Primitive Rebels: Studies in Archaic Forms of Social Movement in the 19th and 20th Centuries, (Norton, New York, 1`959)

Hobsbawm, Eric, The Age of Revolution: 1789-1848, (Mentor, New York, 1962)

Hobsbawm, Eric, Industry and Empire: From 1750 to Present Day, (Penguin, Harmondsworth, Middlesex, 1968)

Hochschild, Arlie, Russell, Strangers in their Own Land: Anger and Mourning on the American Right, (New Press, New York, 2016)

Hollinger, David, "T.S. Kuhn's A Theory of Science and Its Implications for History," American Historical Review, (78/2, April, 1973), p. 370-393.

Homans, George C., English Villagers of the Thirteenth Century, (Norton, New York, 1941)

Hrdy, Sarah Blaffer, Mothers and Others: The Evolutionary Origins of Mutual Understanding, (Harvard University Press, Cambridge Massachusetts, 2009)

Jaspers, Karl, The Origin and Gola of History, (Routledge, [1953], 2001)

Kantorowicz, Ernst, The King's Two Bodies: A Study in Medieval Political Theology, (Princeton University Press, Princeton, 1957)

Khalidi, Rashid, Resurrecting Empire: Western Footprints and America's Path in the Middle East, (Beacon, Boston, 2004)

Klein, Richard G., The Human Career: Human Biological and Cultural Origins, (University of Chicago Press, Chicago, 1999)

Koditschek, Theodore, Class Formation and Urban Industrial Society: Bradford, 1750-1850, (Cambridge University Press, Cambridge, 1990)

Koditschek, Theodore, "Marxism and the Historiography of Modern Britain: From Engels to Thompson to Deconstruction and Beyond," in Terry Brotherstone and Geoff Pilling eds., History, Economic History and the Future of Marxism, (Porcupine Press, London, 1996)

Koditschek, Theodore, "The Possibilities of Theory: Thompson's Marxist History," in Roger Fieldhouse and Richard Taylor eds., E.P. Thompson and English Radicalism, (Manchester University Press, Manchester, 2013)

Koditschek, Theodore, Race and Liberal Imperial Capitalism: The British World Order, 1830-1914, (Under contract to Cambridge University Press)

Koditschek, Theodore, "An Evolutionary Approach to Complex Hierarchical Societies," Behavioural Processes, 161 (2019) 117-28

Krakauer, David C., Gaddis, John Lewis, Pomeranz, Kenneth, eds., History, Big History, and Metahistory, (Santa Fé Institute Press, Santa Fé, 2017)

Kuhn, Thomas, The Structure of Scientific Revolutions, (University of Chicago Press, Chicago, 1962)

Laland, Kevin, and Brown, Gillian, Sense and Nonsense: Evolutionary Perspectives on Human Behaviour, (Oxford University Press, Oxford, 2002)

Laland, Kevin, Darwin's Unfinished Revolution: How Culture made the Human Mind, (Princeton University Press, Princeton, 2017)

Lears, Jackson, Fables of Abundance: A Cultural History of Advertising in America, (Basic Books, New York, 1994)

Le Roy Ladurie, Emmanuel, Montaillou: The Promised Land of Error, (Brasillier, New York, 1978).

Lewens, Tim, Cultural Evolution, (Oxford University Press, Oxford, 2015)

Lewontin, Richard, The Triple Helix: Gene, Organism and Environment, (Harvard University Press, Cambridge Massachusetts, 2000)

Lewontin, Richard, "Does Culture Evolve," (Lecture at Santa Fé Institute, 2012, https://www.youtube.com/watch?v=OEUoqmeO5hs)

Maddison, Angus, Contours of the World Economy: 1-2030 AD, (Oxford University Press, Oxford, 2007)

Maddison, Angus, The World Economy, 2 vols., (Organisation for Economic Cooperation and Development, Paris, 2006)

Marx, Karl and Engels, Friedrich, The Communist Manifesto, (International Publishers, [1848], 1948)

Marx, Karl, Capital, vol. I, (Progress Publishers, Moscow, [1867], 1954)

Mathias, Peter, The First Industrial Nation, (Methuen, London, 1983)

Mayer, Arno, J, Dynamics of Counterrevoution in Europe, 1870-1956: An Analytical Framework, (Harper, New York, 1971)

Maynard Smith, John, The Theory of Evolution, (Penguin, Harmondsworth, Middlesex, 1975

Maynard Smith, John and Szathmàry, Eörs, The Major Transitions in Evolution, (Oxford University Press, Oxford, 1995)

Maynard Smith, John and Szathmàry, Eörs, The Origins of Life: From the Birth of Life to the Origins of Language, (Oxford University Press, Oxford, 1998)

McNeill, J.R.and McNeill, William H., The Human Web: A Bird's Eye View of World History, (Norton, New York, 2003)

McNeill, William H., The Pursuit of Power: Technology, Armed Force and Society since AD 1000, (University of Chicago Press, Chicago, 1982)

McNeill, William H., The Rise of the West: A History of the Human Community, (University of Chicago Press, Chicago, 1963).

McNeill, William, Plagues and Peoples, (Doubleday, New York, 1977)

Merton, Robert K., Social Theory and Social Structure, (Free Press, Toronto, 1964)

Mesoudi, Alex, Cultural Evolution, (University of Chicago Press, Chicago, 2011)

Miller, John H., and Page, Scott E., Complex Adaptive Systems: An Introduction to Computational Models of Social Life, (Princeton University Press, Princeton, 2007)

Mitchell, B.R., Abstract of British Historical Statistics, (Cambridge University Press, Cambridge, 1962)

Mokyr, Joel, The Lever of Riches: Technological Creativity and Economic Progress, (Oxford University Press, Oxford, 1990)

Mokyr, Joel, The Enlightened Economy: An Economic History of Britain, (Yale University Press, New Haven, 2009)

Montague, Ashley, ed., Sociobiology Examined, (Oxford University Press, Oxford, 1980)

Moore, Barrington, Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World, (Beacon, Boston, 1966)

Morris, Ian, The Measure of Civilization, : How Social Development Shapes the Fate of Nations, (Princeton University Press, Princeton, 2013)

Odling-Smee, John, Laland, Kevin, and Feldman, Marcus, Niche Construction: The Neglected Process in Evolution, (Princeton University Press, Princeton, 2003)

Ostrom, Elinor, Governing the Commons, (Cambridge University Press, Cambridge, 2015)

Palmer, R.R, The Age of Democratic Revolutions, (Princeton University Press, Princeton, 1959, 1964)

Parker, Geoffrey, Global Crisis: War, Climate Change and Catastrophe in the Seventeenth Century, (Yale University Press, New Haven, 2013)

Parthasrathi, Prasannan, Why Europe Grew Rich and Asia Did Not: Global Economic Divergence, 1600-1850, (Cambridge University Press, 2011)

Perry, Nick, "A Comparative Analysis of 'Paradigm' Proliferation," British Journal of Sociology, (28/1, March 1977), p. 38-50

Pines, David, ed., Emerging Syntheses in Science: Proceedings of the Founding Workshops of the Santa Fé Institute, (Santa Fé Institute Press, Santa Fé, 2019)

Pirenne, Henri, Medieval Cities, Their Origins and the Revival of Trade, (Princeton University Press, Princeton, 1952)

Polanyi, Karl, The Great Transformation: The Political and Economic Origins of our Time, (Beacon Press, Boston, 1944)

Pomeranz, Kenneth, The Great Divergence: China, Europe and the Making of the Modern World Economy, (Princeton University Press, 2000)

Postman, Neil, Amusing Ourselves to Death: Public Discourse in the Age of Show Business, (Penguin, New York, 1985)

Richards, Robert, Darwin and the Emergence of Evolutionary Theories of Mind and Behavior, (University of Chicago Press, Chicago, 1987)

Richerson, Peter and Boyd, Robert, Not by Genes Alone: How Culture Transformed Human Evolution, (University of Chicago Press, Chicago, 2005)

Rodney, Walter, How Europe Underdeveloped Africa, (Howard University Press, Washington, 1982)

Rostow, W.W., ed., The Economics of Take-off into Sustained Growth, (London, 1963)

Rudé, George, Revolutionary Europe: 1783-1815, (Harper, New York, 1964)

Runciman, W.G.A., The Theory of Cultural and Social Selection, (Cambridge University Press, Cambridge, 2009)

Ryan, Mary, Cradle of the Middle Class: The Family in Oneida County, New York, 1790-1865, (Cambridge University Press, Cambridge, 1981)

Sahlins, Marshall, The Use and Abuse of Biology: An Anthropological Critique of Sociobiology, (University of Michigan Press, Ann Arbor, 1976)

Scott, James C., Against the Grain: A Deep History of the Earliest States, (Yale University Press, New Haven, 2017).

Segerstråle, Ullica, Defenders of the Truth: The Sociobiology Debate, (Oxford University Press, Oxford, 2000).

Sewell, William H., Capitalism and the Emergence of Civic Equality in Eighteenth Century France, (Cambridge University Press, Cambridge, 2021)

Sewell, William, Logics of History: Social Theory and Social Transformation, (University of Chicago Press, Chicago, 2005)

Sklar, Katherine, Catherine Beecher: A Study in American Domesticity, (Norton, New Yor, 1976)

Sober, Elliott and Wilson, David Sloan, Unto Others: The Evolution and Psychology of Unselfish Behavior, (Harvard University Press, Cambridge Massachusetts, 1998)

Sperber, Dan, Explaining Culture: A Naturalistic Approach, (Blackwell, Oxford, 1996)

Stone, Lawrence, The Causes of the English Revolution: 1529-1642, (Harper, New York, 1972)

Tawney, R.H., Religion and the Rise of Capitalism, Harcourt Brace Jovanovich, New York, 1926)

Thompson, E.P., The Making of the English Working Class, (Vintage, New York, 1963)

Tilly, Charles, Popular Contention in Great Britain: 1758-1834, (Harvard University Press, Cambridge, Massachusetts, 1995

Turchin, Peter, Historical Dynamics: Why States Rise and Fall, (Princeton University Press, Princeton, 2003)

Von Tunzelman, G.N., Steam Power and British Industrialization to 1860, (Oxford University Press, Oxford, 1978)

Wallerstein, Immanuel, The Modern World System, 3 vols, (Academic Press, New York, 1976, 1980, 1989)

Walzer, Michael, The Revolution of the Saints: A Study in the Origins of Radical Politics, (Atheneum, New York, 1965)

Weber, Max, The Protestant Ethic and. the Spirit of Capitalism, (Scribners, New York, 1958)

Williams, G.C., Adaptation and Natural Selection, (Princeton University Press, Princeton, 1966)

Wilson, Charles, The Dutch Republic, (McGraw Hill, New York, 1968)

Wilson, David Sloan, Darwin’s Cathedral: Evolution, Religion, and the Nature of Society, (University of Chicago Press, Chicago, 2002)

Wilson, E.O., On Human Nature, (Harvard University Press, Cambridge, Massachusetts, 1978)

Wilson, E.O., Sociobiology: The New Synthesis, (Harvard University Press, Cambridge, Massachusetts, 1975)

Wolf, Eric R., Peasant Wars of the Twentieth Century, (Harper, New York, 1969)

Wolf, Eric R., Europe and the People Without History, (University of California Press, Berkeley, 1982)

Part III: Why should we be skeptical of generalizations of Darwinism?

Chapter 12

Is natural selection physical?

Sylvain Charlat[i], Thomas Heams[ii] & Olivier Rivoire[iii]

Abstract

Biology, in contrast to other historical disciplines such as cosmology or geology, is not explicitly articulated with physics. More specifically, its unifying principle, evolution by natural selection, is currently not formulated in physical terms. This hinders any attempt to explore whether this principle may apply to other physical systems, beyond life as we know it, or to understand the origin of life in a physico-chemical framework. To better understand whether an explicit articulation is achievable, we first aim to clarify, on the basis of examples, how principles are articulated within the physical sciences, or between the physical sciences and other scientific fields. This leads us to establish a typology where we emphasize that physical principles involve both "rules" in the form of mathematical relationships between concepts, and "premises", defining the conditions and objects to which they apply; articulations may take place at these two levels. We then ask whether the principle of evolution by natural selection may fit in such a typology of articulations. We contend that addressing this question is made difficult by an apparent but ineffective distinction between rule and premises in current accounts of the principle of natural selection. These reduce evolution by natural selection to the iteration of a constant rule, thus failing to recognize that biological evolution is a process that recursively modifies its own modes of operation, e.g., through changes in inheritance systems or levels of individuality. While this may be ignored when focusing on paradigmatic cases of natural selection (as formalized by population genetics, where connections with physics are recognized), it becomes a patent problem in more general formulations of natural selection. We conclude by discussing whether this problem could be resolved, through a formal and general description of this principle, where rules and premises would be truly independent or,

alternatively, whether its heuristic value, within biology or beyond, is just of a different nature than that of physical principles.

---

[i] Laboratoire de Biométrie et Biologie Évolutive, Université de Lyon, Université Lyon 1, CNRS, UMR 5558, 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne, France. Email : sylvain.charlat@cnrs.fr.

[ii] INRAE, UMR 1313, Génétique Animale et Biologie Intégrative, Domaine de Vilvert Bâtiment 211, 78352 Jouy-en-Josas, France. Email : thomas.heams@agroparistech.fr.

[iii] Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, Université PSL, 75005 Paris, France. Email : olivier.rivoire@college-de-france.fr.

12.1 Introduction

Insofar as living beings are recognized as physical objects, the principle of evolution by natural selection must be physical in some sense. Yet, despite its 160 years of existence, it has not become part of the physicists' toolbox, which makes it non-physical in practice. Going beyond these two obvious but contradictory assertions is the objective of the present essay: trying to clarify in what sense (if any) may natural selection be seen as physical, with at least three underlying motivations. One is to assess the possibility of applying this principle to physical yet non-biological systems (Charlat et al., 2021), that is, beyond living beings and their derivatives, from languages to computer programs. A second, related, motivation is to clarify whether natural selection, in its present formulation, may be appropriate to understand the continuous transition from inanimate to living matter. A third motivation, stemming from a physicist's perspective, is to put it on par with other physical principles. These endeavors would greatly benefit from an explicit articulation of the evolutionary theory with physics and, reciprocally, may be substantially hindered if such an articulation turns out to be unachievable.

Our analysis begins with a survey of the various means by which principles are articulated within physics, or between physics and other scientific fields. We emphasize that physical principles involve "rules" (analogous to mathematical functions) as well "premises" (defining their conditions of applications) and that articulations may take place at both levels. We then discuss whether the principle of natural selection may fit in this typology of articulations. Our analysis suggests that, to some extent, this principle is already articulated with physics, for instance through shared mathematical concepts, notably in its well circumscribed and formalized version developed within population genetics. However, current accounts, when formulated in terms of rules and premises, face a fundamental limitation: the phenomenon of biological evolution inevitably provides examples where the "rules" are themselves evolving and thus indistinguishable from the "premises", as previously emphasized (Godfrey-Smith, 2009). It thus does not seem reducible to a standard recursive function, that would remain constant across time steps. In the concluding section, we discuss whether such a difficulty could be resolved and to what extent it impedes the search for evolution by natural selection in other physical systems.

12.2 What is physical

Considering how concepts and principles are articulated within physics, the first articulation to come to mind is one by derivation, where a principle is explained as an application of more general principles. A textbook example is Kepler's laws of planetary motions, that were explained by Newton as a consequence of his law of gravitation and his laws of motion. This may be symbolically represented as $k = m \circ g$, where a principle is written here as a function $f: P \to I$, from a set of premises P to a set of implications I, and where the premises can be instantiated by $x \in P$ to lead to predictions $f(x)$. The notation $k = m \circ g$ refers to function composition, i.e., $m \circ g(x) = m(g(x))$, where k represents Kepler's laws, m represents the laws of motion and g the law of gravitation. Another example would be the derivation of the classical laws of motion from special relativity in the limit where velocities are small compared to the speed of light.

Many physical principles are, however, irreducible even though they concern emerging phenomena whose constituents are fully described by lower-level principles (Anderson, 1972). Many examples can be found in the field of condensed matter physics; thus, the absence of a critical point on the melting curve of any substance is explained by the impossibility to change symmetry gradually, a basic principle that is not derivable from other physical principles. This principle can be instantiated: the liquid phase is isotropic while the solid phase has the discrete symmetry of a crystal, and this symmetry can itself be derived from properties of the constituent molecules. Symbolically, the fundamental principle $f: P \to I$ is articulated to lower-level concepts $C \in P$ or/and to lower-level principles $g: Q \to P$ such that we may consider $f(x)$ for $x \in C$ or $f \circ g(y)$ for $y \in Q$.

Notably, we find these two types of articulation not only among physics-born principles, but also when considering how a principle originating from outside physics has become articulated with physical ones. An example is information theory, which we understand here in its broadest sense, as the study of phenomena involving the transmission, processing, extraction, and utilization of information. Information theory includes mathematically well formulated principles among which Shannon's theorems, which set fundamental limits to the rate at which data can be compressed and communicated (Shannon, 1948). These theorems have a status analogous to that of fundamental principles of condensed matter physics (Anderson, 1972): they stand on their own and are not reducible to other physical principles, but their premises can be instantiated with physical concepts that are themselves subject to physical principles. For instance, bits can be realized with magnetic materials and their processing is subject to Shannon's theorems. The other type of articulation, by derivation, has also been proposed by considering that physical principles may follow from more general informational principles rather than the opposite. A well formulated case is Jayne's derivation of statistical mechanics from a principle of statistical inference (Jaynes, 1957) and a more speculative one is Wheeler's proposal to derive 'It from Bit' (Wheeler, 1989).

The most fruitful articulations between information theory and physics are, however, of different natures. The major one is the formal articulation between Shannon's theorems and statistical physics, coming from their common reliance on asymptotic principles (law of large numbers). Symbolically, this articulation can be represented as $f = l \circ g$ and $\varphi = l \circ h$ where f is an information theoretic principle, φ a physical principle and l a common underlying principle. In practice, this implies that the two fields share common methods and common concepts, for instance the same concept of entropy.

Finally, another kind of articulation, also exemplified with information theory, is more conceptual. The concept of information appeared in physics first informally, in Maxwell's thought experiment of a demon violating the second law of thermodynamics (Leff & Rex, 2003). The resolution of this paradox involved recognizing which information processing steps are subject to physical constraints, i.e., recognizing which concepts C from information theory were subject to a physical law $\varphi: C \to I$. A solution is provided by Landauer's principle which establishes an equivalence between logically irreversible operations (e.g., data erasure) and thermodynamical irreversible operations (dissipative processes). Another example of a formal articulation between physics and information theory is the development of the field of quantum information (Nielsen & Chuang, 2010), which now finds instantiation in the engineering of quantum computers.

To sum up, the relationships between physics and information theory illustrate four types of articulations that we may divide in two classes. Starting from a well formulated mathematical principle, we may have the first type, *articulation by derivation*. Noteworthily, this can go from physics to another field but also the other way round (e.g. Jaynes' derivation of statistical mechanics from the principle of maximum entropy). Within this first class, we also have a second type, which we call *formal articulation* (e.g. the common asymptotic principles behind Shannon's theorem and thermodynamics). Alternatively, a second class of articulations starts from a concept that may or may not be formalized (i.e., may or may not be the premise of a mathematically formulated principle), which includes the third type, *articulation by* *instantiation* (e.g. the application of Shannon's theorem to physical information processing systems) and the fourth type, *conceptual articulation*, involving the formulation of a new principle (e.g., Landauer's principle).

Before considering which of these four kinds of articulation(s) may be relevant to describe the relation between the principle of evolution by natural selection and physics, it is also worth noting that an informal concept, even if it originates from physics, may find no clear articulation with physical principles; in that sense, it may be considered as 'non-physical'. For instance, the concept of dissipative structure was proposed to explain a broad range of far-from-equilibrium systems exhibiting spatial or/and temporal patterns, including biological evolution (Prigogine, 1969). It has however been shown that no general principle (technically, no variational principle) can cover all these phenomena (Landauer, 1975). This does not mean that no physical prediction can be made by analyzing a particular phenomenon representing a dissipative structure, but that no new prediction can be made from recognizing that this physical phenomenon is an instantiation of the concept of dissipative structures. In other words, dissipative structures can be regarded as "non-physical" since they are not the premise of any physical principle. This example illustrates again that we are taking the question "is *x* physical ?" in an epistemic sense, without questioning the materiality of the entities at play. It also illustrates that articulation by instantiation in absence of a rule is not sufficient to make a concept physical.

12.3 The case of the evolutionary theory

To discuss if and how the principle of evolution by natural selection may be articulated with physics under the above-described typology, it is first necessary to review how it is usually formalized. One of the most cited formulation takes the form of premises, through a list of necessary conditions for evolution by natural selection, as given by Lewontin (Lewontin, 1970) and many subsequent authors (e.g. (Godfrey-Smith, 2009)). These may be hierarchically organized as illustrated in the upper part of figure 1. First, "populations" are required: evolution by natural selection does not apply to individual entities, but to collections of such entities. Second, these populations must be heterogeneous, i.e., harbor some variations in properties that are often denoted as "traits". These variable traits must further fulfill two conditions: (1) be somewhat stable over time, or heritable in systems where reproduction takes place, and (2) affect the stability or the reproductive success of their carriers (their “fitness”).

One the other hand, another common formalization, the Price equation, is more akin to a rule (lower part of figure 1). In contrast to many models from populations genetics, which may also be taken as rules, the Price equation appears most general, not relying on restrictive assumptions such as a particular mechanism of inheritance (Frank, 2012; Gardner, 2020; Luque, 2017; Price, 1970). This equation simply expresses the change in mean value of a trait between two time points as resulting from the “co-variance between the trait and fitness”, but not only so if the trait value also changes at the individual level (that is, if the trait is not perfectly heritable). In Steven Frank’s words (Frank, 2018) : “*The abstract Price equation describes dynamics as the change between two sets. One component of dynamics expresses the change in the frequency of things, holding constant the values associated with things. The other component of dynamics expresses the change in the values of things, holding constant the frequency of things*”. Through its covariance term, this equation formalizes a “rule” according to which the above-defined premises should produce change over time in the population mean of a trait value.

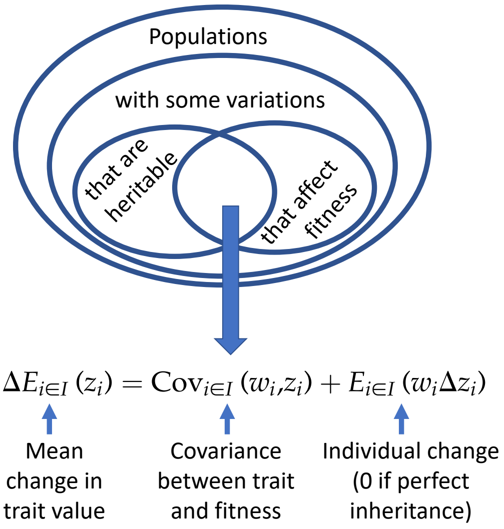


Figure 1. Evolution by natural selection, in its currently most general formulation. The upper part is a hierarchically organized list of premises. The lower part is the Price equation, the rule according to which those premises give rise to evolution, that is, to a change in the mean value of any trait.

Following the above-established typology, let us now try to clarify how evolution could be articulated with physics, starting with the possibility of an articulation by derivation. Strictly speaking, the proposal of deriving the Price equation from a more general principle is meaningless, because this equation happens to be a mathematical identity (Frank, 1995). Yet we note that, independently of the Price equation, multiple proposals have been made to express evolution by natural selection in a physical framework, (e.g. (Bernstein et al., 1983; Lotka, 1922)), although none has been conclusive. Notably, Prigogine and co-workers proposed to view evolution as a particular instance of dissipative structures (Prigogine, 1969) but, as noted above, no general principle applies to dissipative structure, so that evolution cannot be meaningfully reduced to this concept.

A second possibility is that of a formal articulation, where common principles would be recognized as underlying natural selection and physical principles. Here again multiple proposals have been made. Some work follows the goal of identifying common underlying mathematical principles behind Price equation and physical laws (Frank, 2018). Several formal mappings have been found between models of populations genetics and models of statistical physics, which follow from common mathematical principles (Barton & Coe, 2009). These mappings, however, are only established for specific models of population genetics, that is, to formal accounts of particular cases of evolution by natural selection, that may be considered as "paradigmatic" (Godfrey-Smith, 2009).

A third possibility, that we now discuss in more details, is that of an articulation by instantiation, where a physical realization of the premises is formulated: are Lewontin's conditions amenable to physical implementations? A potential problem in addressing this question is that evolution by natural selection, as it can be currently witnessed in biology, applies to objects (traits within individuals within populations,etc…) that are also its *products* : owing to its multigenerational component, evolution is recursive. In principle, this should not constitute a fundamental impediment to its articulation by instantiation with physics: recursive processes may be well formalized through recursive mathematical functions. Yet, we encounter several difficulties when trying to formalize evolution by natural selection along those lines.

A first difficulty comes from the 'dynamical insufficiency' of the Price equation: formally, it cannot be iterated because it requires in its premises more than it delivers in its conclusions (it requires a covariance and delivers only a change in mean trait) (Frank, 1995; Lewontin, 1974). More circumscribed models in populations genetics avoid this caveat, but as noted above, they cannot be taken as general descriptions of natural selection. A second difficulty relates to the fact that recursive functions require a starting point to be effectively iterated: objects that satisfy the conditions but *are not* the products of evolution. One may think for example of clay crystals (Bedau, 1991). However, Lewontin's conditions are at best loosely met in such systems where, in particular, a clear description of individuality is lacking. Computer programs or polymers subject to in vitro evolution can be seen as more satisfactory candidates: they can clearly be formulated in physical terms only, despite being themselves a product of evolution by natural selection. A third and major difficulty comes from acknowledging that no formalization of the principle of natural selection has yet been proposed where it cannot be argued that the rule itself may be subject to change by natural selection. For example, inheritance systems or levels of individuality can be considered as fixed in the short term and part of the rule but are also subject to evolution in the long run. A similar argument has led Goldenfeld and Woese to propose that evolution is "self-referential" (Goldenfeld & Woese, 2011).

In fact, the above listed first and second difficulties may be symptoms stemming from this more general problem: current accounts of evolution cannot be formally framed as rules and premises, because they fail to capture that the plasticity of the phenomenon of biological evolution, where examples are always found where the rules themselves are evolving. No formalization is currently available of a general principle that would apply to the diversity of forms that evolution by natural selection can take. The view of evolution as happening in populations of well-defined individuals harboring well defined traits (that underlies Lewontin's formulation or Price equation) is in fact an idealized account of an end-product of evolution, which is to be explained as much as it is an explanation. This conclusion relates to the previously emphasized argument that even within the biological world, many border-line cases (as opposed to "paradigmatic" ones) can be found, where this framework does not apply straightforwardly (Godfrey-Smith, 2009).

Finally, let us consider more briefly the fourth possibility, that of a conceptual articulation of natural selection with physics, where informal concepts are formalized and shown to be subject to physical principles. In fact, many physical principles have been formulated to apply to biological systems, constituting the field of biophysics; but this discipline tends not to refer to evolution. Several recent works in stochastic thermodynamics may be seen as filling this gap, including for instance efforts to identify thermodynamic limits to replication (England, 2013). More broadly, biological evolution has long been an important source of inspiration in physics and engineering. Current work on functional, “adaptive” or even “intelligent” matter, which can modify its internal structure in response to external stimuli from the environment (Kaspar et al., 2021)￼ may thus be expected to unravel new physical principles pertaining to the evolutionary notions of function and adaptation.

12.4 Perspectives

Our analysis suggests that some articulations are already effective between evolutionary theory and physics. A formal articulation takes place when common underlying principles are shared, which permits methods to be transferred between physics and population genetics (Barton & Coe, 2009), a particular branch of evolutionary biology, grounded in a particular inheritance system, where the “rules” are regarded as constant. A conceptual articulation is also effective when concepts originating from evolutionary biology are inspiring new physics (e.g. England, 2013; Kaspar et al., 2021). However, no articulation by derivation has been achieved, where the principle of natural selection would follow from more elementary and general physical principles (or reciprocally). This is not unexpected, considering that even within physics, many emerging principles are irreducible. Maybe more surprisingly, even an articulation by instantiation, whereby the premises of evolution would be formulated in physical terms, encounters difficulties. In our view, this arises from intricacies between the rule and the premises: evolution not only applies to its own products, which may be captured by a recursive mathematical function, but also changes its own rules of operation, like a recursive function that would change itself across time steps. In other words, given a precisely defined rule, e.g. a population genetics model, we can find examples in biological evolution where elements of the rule are themselves considered as subject to natural selection.

Could this problem be resolved? We can at least speculate on what its solution would look like. One possibility would be to stick to the rule / premises framework but noting that the rule of natural selection should be a “meta-rule”, a rule-changer, describing how modes of evolution by natural selection are evolving themselves, through changes of features such as inheritance systems, rates and modes of mutation, or levels of individuality. Another possibility would be to recognize that a satisfactory description of the evolutionary process may take a radically different form. As previously argued (Goldenfeld & Woese, 2011), we may even need different mathematical concepts to formalize evolution *in general* (Fontana & Buss, 1994), which may in turn suggest new modes of articulations with physics.

It may also be that natural selection *in general* cannot be mathematically formalized, just as dissipative structures cannot be associated with a unifying principle. This would arguably hinder the search for natural selection beyond life, as well as the integration of natural selection as an essential component in the physico-chemical emergence of "lifeness". But would it necessarily imply that natural selection cannot be of any heuristic value outside of its original field? Within biology, natural selection serves as a general and often implicit explanation for adaptations, and thus as a justification for "functional thinking": the heuristic assumption that many features of biological systems are *best understood* as fulfilling roles within complex ensembles that constitute a living whole, the individual, the organism. Here, *best understood* means that capturing the function of a feature provides a mean to summarize its important properties, its "evolutionary causes", without focusing on unnecessary details: a wing is a feature that allows flying, regardless of what molecules it is made of. This type of reasoning is reminiscent of the application of variational principles to describe physical phenomena. For instance, the laws of refraction (a local property) can be derived from a principle of least action, namely the extremization of the time taken by light to join two points (a global property), or the equilibrium states of matter can be derived from the minimization of an appropriate thermodynamic potential. This has inspired past attempts to derive a general physical principle related to that of natural selection, as typically illustrated by works on dissipative structures, but so far to no avail. While an explicit articulation of natural selection with physics may still be sought along those lines it remains possible in the meantime to explore whether the particular kind of explanations it provides to biologists could be relevant elsewhere.

References

Anderson, P. W. (1972). More Is Different: Broken symmetry and the nature of the hierarchical structure of science. *Science*, *177*(4047), 393–396. https://doi.org/10.1126/science.177.4047.393

Barton, N. H., & Coe, J. B. (2009). On the application of statistical physics to evolutionary biology. *Journal of Theoretical Biology*, *259*(2), 317–324. https://doi.org/10.1016/j.jtbi.2009.03.019

Bedau, M. (1991). Can Biological Teleology Be Naturalized? *Journal of Philosophy*, *88*(11), 647–655. https://doi.org/10.5840/jphil1991881111

Bernstein, H., Byerly, H. C., Hopf, F. A., Michod, R. A., & Vemulapalli, G. K. (1983). The Darwinian Dynamic. *The Quarterly Review of Biology*, *58*(2), 185–207. https://doi.org/10.1086/413216

Charlat, S., Ariew, A., Bourrat, P., Ferreira Ruiz, M., Heams, T., Huneman, P., Krishna, S., Lachmann, M., Lartillot, N., Le Sergeant d'Hendecourt, L., Malaterre, C., Nghe, P., Rajon, E., Rivoire, O., Smerlak, M., & Zeravcic, Z. (2021). Natural selection beyond life? A workshop report. *Preprints*, *11*, 1051. https://doi.org/DOI:

England, J. L. (2013). Statistical physics of self-replication. *The Journal of Chemical Physics*, *139*(12), 121923. https://doi.org/10.1063/1.4818538

Fontana, W., & Buss, L. (1994). "The arrival of the fittest": Toward a theory of biological organization. *Bulletin of Mathematical Biology*, *56*(1), 1–64. https://doi.org/10.1016/S0092-8240(05)80205-8

Frank, S. A. (1995). George Price's contributions to evolutionary genetics. *Journal of Theoretical Biology*, *175*(3), 373–388. https://doi.org/10.1006/jtbi.1995.0148

Frank, S. A. (2012). Natural selection. IV. The Price equation*: Price equation. *Journal of Evolutionary Biology*, *25*(6), 1002–1019. https://doi.org/10.1111/j.1420-9101.2012.02498.x

Frank, S. A. (2018). The Price Equation Program: Simple Invariances Unify Population Dynamics, Thermodynamics, Probability, Information and Inference. *Entropy (Basel, Switzerland)*, *20*(12), E978. https://doi.org/10.3390/e20120978

Gardner, A. (2020). Price's equation made clear. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1797), 20190361. https://doi.org/10.1098/rstb.2019.0361

Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. Oxford Univeristy Press.

Goldenfeld, N., & Woese, C. (2011). Life is Physics: Evolution as a Collective Phenomenon Far From Equilibrium. *Annual Review of Condensed Matter Physics*, *2*(1), 375–399. https://doi.org/10.1146/annurev-conmatphys-062910-140509

Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, *106*(4), 620–630. https://doi.org/10.1103/PhysRev.106.620

Kaspar, C., Ravoo, B. J., van der Wiel, W. G., Wegner, S. V., & Pernice, W. H. P. (2021). The rise of intelligent matter. *Nature*, *594*(7863), 345–355. https://doi.org/10.1038/s41586-021-03453-y

Landauer, R. (1975). Inadequacy of entropy and entropy derivatives in characterizing the steady state. *Physical Review A*, *12*(2), 636–638. https://doi.org/10.1103/PhysRevA.12.636

Leff, H. S., & Rex, A. F. (Eds.). (2003). *Maxwell's demon 2: Entropy, classical and quantum information, computing*. Institute of Physics.

Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics*, *1*, 1–18.

Lewontin, R. C. (1974). *The genetic basis of evolutionary change*. Columbia Univ. Pr.

Lotka, A. J. (1922). Natural Selection as a Physical Principle. *Proceedings of the National Academy of Sciences of the United States of America*, *8*(6), 151–154. https://doi.org/10.1073/pnas.8.6.151

Luque, V. J. (2017). One equation to rule them all: A philosophical analysis of the Price equation. *Biology & Philosophy*, *32*(1), 97–125. https://doi.org/10.1007/s10539-016-9538-y

Nielsen, M. A., & Chuang, I. L. (2010). *Quantum computation and quantum information* (10th anniversary ed). Cambridge University Press.

Price, G. R. (1970). Selection and Covariance. *Nature*, *227*(5257), 520–521. https://doi.org/10.1038/227520a0

Prigogine, J. (1969). *Structure, Dissipation and Life, Theoretical Physics and Biology*. North Holland, Amsterdam.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Wheeler, J. A. (1989). Information, Physics, Quantum: The Search for Links. In *Oroceedings III International Symposium on Foundations of Quantum Mechanics* (pp. 354–358).

Chapter 13

The risks of evolutionary explanation

H. Clark Barrett[i]

Abstract

Evolutionary explanations of behavior are special in that they involve both proximate and ultimate components. Proximately, evolutionary accounts posit mechanisms that generate observed patterns of behavior. At the ultimate level, evolutionary accounts explain the existence of these proximate mechanisms via evolutionary processes such as selection or drift acting in the past. Does positing or accepting such explanations carry any risks? Here I consider two kinds of risk, epistemic and ethical. Epistemic risk is the risk of being wrong about a matter of fact, such as whether a posited empirical pattern actually exists, or that a particular set of causes explains it. Ethical risk is the risk of causing harm. If the acceptance of a particular explanation for an observed pattern of behavior entails real-world consequences that could lead to harm, then epistemic risk entails ethical risk. Here I consider these risks in cases of evolutionary explanation, and ask whether evolutionary explanations might carry special risks that we should consider when developing and testing evolutionary hypotheses.

---

[i] Department of Philosophy and Center for Behavior, Evolution and Culture, University of California, Los Angeles.

13.1 Introduction

Evolutionary explanations are a special kind of explanation. Events, traits, behaviors, decisions are always multiply caused, but when we say there is an evolutionary reason for them, we are pointing to a particular kind of cause. This typically has to do with processes of natural selection acting in ancestral environments, or other evolutionary processes such as drift. Such explanations can have a particular weight when it comes to human affairs. Consider, for example, the following passage, from an article in the New York Times Sunday Review in 2014, asking why ADHD—attention deficit hyperactivity disorder—has such a high prevalence in the United States:

> Attention deficit hyperactivity disorder is now the most prevalent psychiatric illness of young people in America, affecting 11 percent of them at some point between the ages of 4 and 17. The rates of both diagnosis and treatment have increased so much in the past decade that you may wonder whether something that affects so many people can really be a disease.
>
> Recent neuroscience research shows that people with A.D.H.D. are actually hard-wired for novelty-seeking — a trait that had, until relatively recently, a distinct evolutionary advantage. Compared with the rest of us, they have sluggish and underfed brain reward circuits, so much of everyday life feels routine and understimulating.
>
> To compensate, they are drawn to new and exciting experiences and get famously impatient and restless with the regimented structure that characterizes our modern world. In short, people with A.D.H.D. may not have a disease, so much as a set of behavioral traits that don't match the expectations of our contemporary culture. (Friedman, 2014)[1]

---

[1] https://www.nytimes.com/2014/11/02/opinion/sunday/a-natural-fix-for-adhd.html

As explanations go, this one feels definitive, weighty, and satisfying. It seems plausible, and it doesn't take much expertise to follow it. There also feels like an evaluative, moral component to it: if ADHD is not a disease, then perhaps we should think differently about people who have it. Some might get the message that their condition is in fact *good* – or at least it was, until very recently. This could provide solace to sufferers and to parents who are worried that there is something wrong with their children. It could make people less, or perhaps more, judgmental or skeptical about others who claim to have ADHD. And the author is a professor of clinical psychiatry, adding authoritative heft to the claim.

But is it true? To address that, we need to go a little deeper into the explanation being offered. There are in fact multiple things being asserted here that together constitute the full evolutionary claim. For example, there is a claim about what evolutionary biologists sometimes call "proximate" causation: that ADHD is "hard-wired." There is also a claim about "ultimate" causation: that the trait was selected for in ancestral environments (supplemented, later, by descriptions of genomic studies with contemporary hunter-gatherer populations—of questionable relevance to the evolutionary explanation of ADHD's current prevalence in the United States). This claim about the ultimate cause of the trait is what makes it an evolutionary explanation.

Together, these claims are ontological: they are statements about the nature of reality. There is also an epistemological claim, or a claim about what we do or don't know: "recent neuroscience research *shows...*" (emphasis mine). While these claims are descriptive—assertions about what's true—the piece also later makes prescriptive suggestions about how ADHD might be treated, and how we might think about it differently, as a potential "asset."

The aim of the present essay is not specifically to debunk the claims in this New York Times article, nor to suggest that we know whether these claims are true or false. The studies and findings described in the article are real, though I have not fact-checked every one of them. However, I think it's also clear that we do not know whether the evolutionary explanation on offer is true or false. Certainly, the evidence described in the article does not show it to be true. And in this case, the evidence doesn't even demonstrate the proximate claim, that ADHD is "hard-wired" (indeed, some biologists question, for good reason, whether any such claim could be true (Elman et al., 1996)).

Instead, what this essay is about is what I am calling the "risks" of evolutionary explanation. First, I will define what I mean by these risks—and how risks that are not themselves about the facts of the case follow from mistakes we might make in taking conjectures for known facts. Second, I will consider a case study from the primary evolutionary literature. Finally, I will conclude with some thoughts about how the incentive structures of academia and popular culture "select for," in Smaldino and McElreath's terms, a low bar in accepting some evolutionary explanations (Smaldino & McElreath, 2016). This places us at risk, both as both producers and consumers of knowledge

13.2Varieties of risk

It is common among some evolutionary researchers, including many evolutionary social scientists, to deny that there are any risks at all associated with evolutionary theorizing. Some of this denial arises as backlash to Gould and Lewontin's depiction of evolutionary explanations, and particularly adaptationist ones, as "just-so stories" that can rarely, if ever, be proven true or false (a position that I, too, have argued is too extreme) (Barrett, 2015; Gould & Lewontin, 1979). Another part of the resistance to thinking about the risks of evolutionary explanation comes from what philosopher Heather Douglas has called the "value-free ideal" of science (Douglas, 2000, 2009). According to this ideal, which many scientists hold, science is only about adjudicating the truth or falsehood of factual claims, and is neither influenced nor should not be concerned with any risks associated with getting the facts wrong other than this risk itself.

This is what we can call *epistemic risk*, which I will define here as the risk of being wrong, but nothing beyond that (Babic, 2019; Pritchard, 2016). When we look at a stopped clock and conclude that it's noon when it's actually 3PM, we are wrong about the facts of the time. In this case there was evidence—the position of the clock's hands—but it misled us. We faced epistemic risk when we trusted the clock as our only source of evidence about the time (the philosophy of science literature is full of advice about when we should or shouldn't trust particular sources of evidence). As a result, we think it's noon but it's actually three hours later: we are wrong, we have formed a false belief about reality, and being wrong about the facts is the downside of epistemic risk.

But, as we all know, there are more risks than just epistemic ones to relying on a faulty clock. Suppose, for example, we had a plane ticket for a flight from Los Angeles to Tokyo at 3PM, which cost us $2000. We were planning on leaving for the airport in L.A. at 1PM, which would have given us plenty of time. But in reality, our flight has left. We have lost $2000, which is unrecoverable. We have to book a new flight. We will miss our friend's wedding, which is early the next morning in Tokyo. In this case, it turns out that our friend takes our missing the flight to be a sign of insufficient care and investment in the relationship, which puts our friendship in jeopardy.

How should we conceptualize these kinds of risks, which hinge on the risks of getting the facts wrong? To my knowledge, there is no good name for this category of risks, because it potentially includes everything. Someone could die if you get the time wrong (e.g., if you are a surgeon). You could die (e.g., if you thought you still had time to take your medication). A nuclear plant could melt down, or a war could start. And there are other ethical and social risks: hurting your friend's feelings, or worse.

Philosophers of science including Carl Hempel, Heather Douglas, and others have pointed out that for reasons like these it is impossible to isolate purely epistemic risk in ways that many scientists naively believe is possible (Douglas, 2000; Hempel, 1954). This is true for both the "inputs" to epistemic risk—the things that raise or lower our probability of being wrong—as well as its outputs, or the downstream consequences of epistemic error.

The first point, about the inputs to epistemic risk, can be illustrated via Hempel's concept of *inductive risk*. Induction refers to inferences from specific cases to general truths. To take a currently relevant example, suppose we want to know whether a particular vaccine is effective against a viral strain, or not. To do so, we need to define some operative level of "effectiveness." This involves choices outside the facts: i.e., there is no objective line at which the proper level of "effectiveness" should be drawn. Any conclusion we reach at any point in time is also, of course, surrounded by some degree of uncertainty, and depends on how much data we gather, the quantity of data, etc.

In this case, we may be able to find out the truth to any level of precision we want, but at some point we have to stop and decide. We need to tell the CDC, for example, to greenlight the vaccine or not. This means that if we have an inductive decision point—"yes, it's effective at preventing 95% of hospitalizations and deaths" or "no, it's not"—we need to make decisions that are inputs to our inductive process and that precede the facts themselves. These decisions involve values other than the facts at hand, such as how much epistemic risk we're willing to accept during an accelerating pandemic. To say that science is about "just the facts" is, in this sense, missing the point that non-epistemic values are required for the entire process to work, all the way through (Douglas, 2009).

It should also be obvious from the examples given above that there are many varieties of risk that are "downstream" from epistemic risk: namely, all the possible practical, real-world consequences of being wrong. In the stopped clock example, I've spelled some of these out. What about the ADHD case? I've alluded to some of the consequences of taking the proposed evolutionary explanation for ADHD to be true, such as shifting peoples' perceptions of whether it is a "disease," or whether it's "good" or "bad." Certainly, taking the trait to be "hard-wired" has implications for what kinds of interventions one might take to precent or treat it. And there are likely many others.

Of the many kinds of risk that we might regard as entailed by epistemic risk—including all kinds of practical, real-world risk—I'd like to consider in particular what we might call *ethical* risks. When we speak of ethical risks we speak, generally, of the risk of harms: physical harms, social harms, emotional harms, and even harms to non-humans, such as harms to the environment. And these risks can be borne by both producers of knowledge, and consumers of it.

As we have seen, anyone can modulate the degree of epistemic risk to which they are exposing themselves by setting higher or lower bars for things like the nature and quantity of evidence and how much evidence they need to accept a conjecture as true. If there are potential harms to epistemic error, then producers and consumers of knowledge expose themselves to ethical risk as a consequence of how high or low a bar they set in accepting a particular explanation as true, or likely to be true. An author who claims that an evolutionary explanation has been "shown" or "proven"—as well as a consumer who accepts this claim—feeds processes of ethical risk. Some of these consequences they might never have considered (and in cases of people who adhere to the "value-free ideal" of science, they might actively deny). It is the nature of these ethical risks, when it comes to evolutionary explanations in particular, that I would like to consider here.

13.3Do evolutionary explanations entail special risks?

It is generally accepted that what makes evolutionary explanations special—what defines them, in fact—is their "ultimate" rather than "proximate" character, in Tinbergen's well-known framework (Tinbergen, 1963). To explain a trait, behavior, event in evolutionary terms, in this framework, is to say that evolutionary processes such as natural selection or drift were part of the causal chain that brought it about, and indeed, were necessary in bring it about. And of course, specific evolutionary explanations, such as the explanation for ADHD offered above, posit specific evolutionary causes: in that case, a specific set of ancestral conditions and a specific functional reason for which the ADHD phenotype would have increased survival and reproduction relative to other variants.

We can ask, as Gould and Lewontin and many others have, whether there are particular epistemic risks associated with evolutionary explanations of this kind. Indeed there are, because a proper evolutionary explanation for a trait such as ADHD must include *both* a proximate explanation (e.g., the trait is "hard-wired") and an ultimate one—so the risk of error must be higher than any explanation that relies on the proximate part alone. And because processes in the past are far more opaque than processes we can observe in the present, the epistemic risk associated with making inferences about those processes is typically higher as well (though not, contrary to what Gould and Lewontin imply, always insurmountable (Currie, 2018)).

Because many kinds of ethical risk are, as explained above, downstream from epistemic risk, this means that the increased epistemic risk of evolutionary explanations must entail, at some level, increased ethical risk as well. That is, to the extent that *any* explanation for something creates some risk of harm, explanations with a higher risk of being wrong (holding all else equal) must lead to higher ethical risk. Since (most) evolutionary explanations have the same explanatory risk as non-evolutionary ones in the proximate part—e.g., in the case of the ADHD hypothesis above, evidence for the "hard-wired" part—requiring the additional, ultimate-level part of the explanation can only increase rather than decrease epistemic risk, and therefore make the risk of any harms associated with getting that explanation wrong greater.

But there is more. It's true that the ultimate component of evolutionary explanations is what makes them "special," but such explanations may carry additional weight due to the role that evolutionary narratives may play inside and outside of science. For example, there is what Lewontin has called the "legitimating" function of evolutionary explanations (Lewontin, 1996).

Unfortunately, traces of this legitimating function can be seen in the discourse about ADHD above. For example, if ADHD has adaptive origins, maybe we shouldn't consider it a "disease." Indeed, we might even see it as an "asset." It's hard to imagine much more legitimating language to describe something.

Any approach to ethics should include things like misunderstandings, misrepresentations, and changes in the valuations of human traits as potential harms. But the potential harms of misunderstanding the causes and nature of a trait like ADHD can extend well beyond mere representational or emotional harm to include things like misdiagnosis or improper treatment. Indeed, there is abundant evidence that how conditions are medicalized and valorized positively or negatively, within and outside the medical community, has huge effects on how doctors prescribe treatment, how patients do or don't seek treatment, and about how they are perceived positively or negatively by themselves and others, affecting a host of real-world "outcomes" from relationship quality to employment and beyond (Kvaale et al., 2013; Rafalovich, 2013). Just this one example—ADHD—shows that there are many possible ethical risks to accepting evolutionary explanations, including many that we might not be aware of when nodding along to a Sunday newspaper article that seems to make a lot of sense.

When it comes to the entanglement between epistemic and ethical risk, many evolutionary scientists point to a principle sometimes known as the "naturalistic fallacy," a term coined by the philosopher G. E. Moore, and related to what is sometimes called the "is-ought" problem originally introduced by Hume (Curry, 2006; Frankena, 1939; Moore, 1959). There are many varieties of this problem, but the way that naturalistic fallacy is often presented by evolutionists is that one cannot make moral inferences about what is right or wrong from facts about what "is," such as whether something is pleasurable, or common in a species, or the product of some evolutionary process.

It is important for students to be exposed to this fallacy (whether it is a fallacy or not, I leave aside for the moment). Unfortunately, however, many students as well as teachers take the naturalistic fallacy to sever the entanglement between epistemic and ethical risk. For reasons explained here, I think that is a mistake. It might be that we need to apply values other than epistemic ones in order to make moral and ethical decisions, and it might be that evolutionary truths, even if we can know them, would give us no moral guidance (this is often the conclusion drawn from discussions of the naturalistic fallacy). But that does not mean that the risk of making false evolutionary conclusions does not entail ethical risk. Indeed it does, regardless of your philosophical position on the naturalistic fallacy.[2]

13.4 The ADHD case

If epistemic risks feed into ethical risks, as I have argued, then it's important to look carefully at reasons why specific evolutionary claims might be wrong. I will briefly consider the case of ADHD before turning to another from the evolutionary literature.

The main claim in the NYT article about ADHD is that "Recent neuroscience research shows that people with A.D.H.D. are actually hard-wired for novelty-seeking — a trait that had, until relatively recently, a distinct evolutionary advantage" (Friedman, 2014). The proximate part of the claim is that "people with A.D.H.D. are actually hard-wired for novelty-seeking," and the ultimate past is that this is "a trait that had, until relatively recently, a distinct evolutionary advantage."

The ”hard-wired” part of the claim is based in part on a study that found that adult participants who had been medically diagnosed with ADHD "had significantly fewer D2 and D3 receptors (two specific subtypes of dopamine receptors) in their reward circuits than did healthy controls"—and similar findings in studies with children are mentioned (Friedman, 2014). The article also suggests that ADHD is linked to having a particular allele of a dopamine receptor gene (DRD4 7R), and mentions a study in a contemporary human population that links variation at this genetic locus to health outcomes (in particular, adult weight).

Readers who have taken introductory genetics and evolutionary biology classes will already see problems with the way these claims are couched in terms of "hard-wiredness," as well as their tenuous nature as sources of evidence for the hypothesis that ADHD is an evolved adaptation, which in turn is proposed to explain why ADHD is so prevalent in the U.S. First,

---

[2] In my experience, presentations of the naturalistic fallacy as a fallacy often go hand-in-hand with endorsements of the value-free ideal of science. The perceived irrelevance of evolutionary facts for moral considerations leaves people free to fire off evolutionary speculations at will, because they believe there can be no harm in such speculations. For reasons presented here, I think this view is mistaken.

any study showing differences in dopamine receptor frequency across two groups of participants does not show that these differences are “hard-wired”—whether the groups are of adults or children, and regardless, essentially, of the cognitive or neural phenotype in question. Brains and neural systems are plastic on the scale of lifetimes, and shorter timescales as well (Bernhardi et al., 2017; Elman et al., 1996). Traits such as the concentrations of membrane proteins are in a state of homeostasis, and are constantly being maintained and adjusted based on local conditions. This means that the frequency of dopamine receptors in an individual may well be a function of short-term or long-term conditions that the individual has experienced. Just showing phenotypic variation in a trait like this does not show that it is “hard-wired,” unless that term is simply being used to say that individual differences in behavior can be linked to differences in the current states of their brains, which is presumably true of *all* behavior. Interestingly (Weisberg et al., 2008) have proposed that neuroscience explanations of this kind have a particularly “seductive allure,” and offer experimental evidence that including neuroscience information in the explanation for a behavior (e.g., that it is linked to a particular brain region) can greatly increase its perceived plausibility.

The use of genomic evidence in this case is problematic as well, for multiple reasons. To see why, consider the evolutionary hypothesis that the evidence is being used to support:

> It was not until we invented agriculture, about 10,000 years ago, that we settled down and started living more sedentary — and boring — lives. As hunters, we had to adapt to an ever-changing environment where the dangers were as unpredictable as our next meal. In such a context, having a rapidly shifting but intense attention span and a taste for novelty would have proved highly advantageous in locating and securing rewards — like a mate and a nice chunk of mastodon. In short, having the profile of what we now call A.D.H.D. would have made you a Paleolithic success story (Friedman, 2014).

The genomic study mentioned in support of this claim compares two Ariaal communities in Kenya, one of which subsists economically on keeping herds of domesticated animals (and who thus move their herds from place to place) and another which subsists on plant agriculture (and who are thus “sedentary”). The study found that in the herders, possession of the ADHD-associated DRD4 receptor is associated with better nourishment and body weight, and the reverse was true in the herders (Eisenberg et al., 2008).

While the findings of this study are interesting and potentially important for understanding how genetic variation impacts variation in health, their use as evidence for the evolutionary claim above is questionable in several ways. For one, showing that the allele leads to higher body weight in herders than farmers does not constitute evidence that this particular allele was selected for in ancestral environments because of its benefits for a hunter-gatherer lifestyle. Contemporary Ariaal people are not ancestors of anyone (except their children), do not live in ancestral environments, and are not hunter-gatherers. Second, the mechanism leading to the genotype-phenotype-environment interaction, here, is unknown. The dopaminergic system influences many things including not just attention but eating behaviors and even digestive functions, meaning that the association needn't be produced by behaviors associated with ADHD (Glavin & Szabo, 1990).

Finally, the link between genomic differences and natural selection is not as clear-cut as this article seems to imply. When there is genomic variation across individuals that correlates with, for example, health differences, that by itself does not suggest that the variation exists *because* of natural selection, nor that particular alleles have experienced long histories of positive selection. Indeed, if the DRD4 7R allele had been under positive selection for millions of years, as the article implies, one would expect relatively little variation at this locus rather than a lot, and for the DRD4 7R allele to be common in contemporary human populations. The evidence that the article offers in support of its evolutionary claim—that 11 percent of American children are diagnosed with ADHD at some point between the ages of 4 and 17—is thus not strong support for the hypotheses of a long history of selection. In fact, it might be seen as evidence against it.[3]

---

[3] One study estimated the global mean frequency of the 7-repeat allele (7R) of DRD4 to be around 20% (Chang et al., 1996).

13.5The seductive allure of evolutionary explanations

Some might argue that basing a case for the ethical risks of evolutionary explanation on a newspaper article is unfair. Such articles are written for a popular audience and are not part of the primary scientific literature. In this case, the article was written by a clinical psychiatrist and based on the primary literature, so the epistemic risk might be seen as depending on the inferences this particular scientist draws from the evidence. However, the fact that this article was written for the public and published in a highly visible and respected platform like the New York Times might actually be seen as increasing, rather than decreasing, the ethical risk of the "ADHD as adaptation" hypothesis. Ethical risks might be amplified by the larger number of consumers of this information, its perceived authority to a lay audience, the ability of pop science writing to persuade by bypassing normal epistemic checks, and the potential appeal of the claim due to the immense popular interest in ADHD among a readership such as that of the New York Times. In a world of information and misinformation, ethical risks are faced not just by producers of information but by consumers as well. We all play a role in spreading ideas because we think they are plausible or worth telling others about. Harmful and false explanations for things, such as the vaccine hypothesis for autism, can spread this way (Del Vicario et al., 2016; Motta et al., 2018; O'Connor & Weatherall, 2019).

As it happens, the phenomenon of affording excessive epistemic credulity to some kinds of evolutionary explanations is not limited to the pop science literature or to experts like physicians who might adopt evolutionary thinking without having formal training in it. Indeed, in evolutionary social science, ultimate, evolutionary explanations are sometimes seen as "better" than merely proximate explanations of the same phenomena.

This belief carries substantial epistemic risk. On the one hand, there is truth to the idea that if we arrive at a *correct* explanation for a psychological, behavioral, or cultural phenomenon that involves both proximate and ultimate components, then this is indeed a more complete explanation of the phenomenon than the proximate explanation alone. But some evolutionary scientists jump from this correct idea to an incorrect one: that when we have on hand one hypothesis that is merely proximate and another hypothesis that involves both proximate and ultimate (evolutionary) components, we should prefer the latter.

As explained above, this is not the case—indeed, the exact opposite is true. From a logical perspective, one can see that a hypothesis, H1, that requires just one thing (A) to be true in order for it to be correct must be a priori less likely than another hypothesis, H2, that requires both that thing (A) *and* another thing (B) to be true. In the ADHD case, this means that the hypothesis that ADHD is "hard-wired" (whatever that might mean, and for whatever reason) is more likely to be true than the hypothesis that ADHD is "hard-wired" because of natural selection, which in turn is more likely than an even more specific hypothesis about the specific sources of selection, such as hunting.

The fact that this is a fairly obvious logical error does not, unfortunately, prevent many scientists from making it. A variety of this error is known as the "conjunction fallacy" – the probability of a conjunction (A&B together) cannot exceed the probabilities of its constituents, A and B—and yet there is experimental evidence that people often commit this fallacy (Tversky & Kahneman, 1983). For some evolutionary social scientists, the special appeal of ultimate, evolutionary explanations seems to push them especially hard to make this mistake, at least in some cases. On analogy to Weisberg et al.'s notion of the "seductive allure of neuroscience explanations," we can call this the "seductive allure of evolutionary explanations" (Weisberg et al., 2008).

13.6Life history theory as a case study

Let me give an example—or at least, a potential example—from the primary evolutionary literature: explanations of individual differences based in Life History Theory (LHT). I call this a "potential" example because I do not in fact know whether the evolutionary hypotheses in question are correct or incorrect, nor whether evolutionary researchers are actually favoring them compared to merely proximate versions, and / or versions that deny the ultimate part of the explanation (i.e., (A & not B)). I strongly suspect that some evolutionary researchers are making this mistake, but I will present no evidence for that other than my own intuitions. I will, however, argue that acceptance of or preference for evolutionary explanations over and above merely proximate explanations, or explanations that hold the specific evolutionary mechanisms proposed are *not* at work, carries epistemic risk and subsequent ethical risk.

In the case of LHT, evolutionary explanations entail evolved mechanisms which are posited to be the proximate causes of observed developmental patterns and individual differences in psychology and behavior. There are several kinds of epistemic risk here. First, we might be wrong about the patterns of behavior we are trying to explain. This might be a particular risk in cases where the data comes from experimental psychology studies, as much of the data in LHT studies in evolutionary psychology do (Simmons et al., 2011). Second, we might be wrong about the proximate mechanisms—the ones posited to have particular evolutionary origins—that are being used to explain these observed patterns. Third, we might be wrong about the ultimate reasons for existence of those proximate mechanisms. In principle, we could be wrong about (3) even if we are right about (1) or (2).

All of these epistemic risks are causes for concern, but here I'd like to focus mostly on risk (2): the risk of being wrong about the proximate mechanisms leading to observed individual differences in behavior and psychology. I'm also concerned with the interplay between risks (2) and (1), since the data used to support or hypotheses about evolved mechanisms are also generated specifically to test those hypotheses, and therefore can be influenced in certain ways by them (e.g., designed in ways favorable to the hypotheses).

For life history theory (LHT), I am specifically concerned with "within-species" varieties of life history that posit mechanisms of facultative developmental calibration to the environment, within individuals of the same species—in this case, humans (Stearns & Rodrigues, 2020; Zietsch & Sidari, 2019). There are lots of theoretical details and nuances here, but what I am particularly interested in are hypotheses about mechanisms that speed up or slow down developmental timing, and adjust a host of psychological traits, as a function of growing up in environments that bode well or poorly for a long lifespan.

Some evolutionary theorists have suggested that natural selection may shape developmental mechanisms to facultatively adjust developmental timing based on environmental cues of long or short life expectancy. According to this theory, when a child grows up in an environment with cues to short life expectancy, their developmental timing should be sped up, so that they reach developmental milestones such as puberty earlier and reproduce sooner than they otherwise would have. This is because the costs of early reproduction are outweighed by the benefits of getting reproduction done within the briefer time window available. Along with this, LHT theorists have posited a suite of psychological characteristics that should be developmentally calibrated by this mechanism: for example, low life expectancy cues should trigger a psychology of greater risk-taking and shorter psychological time horizons—a kind of "live fast, die young" psychology. On the other hand, children who grow up in cushier, safer environments with longer life expectancy are expected to exhibit greater patience, less impulsivity, increased planning and deliberation about the future, and waiting longer to accumulate wealth and have children (Copping et al., 2013).

This is sometimes called the “fast-slow” life history continuum. The framework was originally developed to explain differences in species-typical life history traits across taxa (e.g., the life history differences between elephants and mice). The variety that much work in evolutionary psychology has focused on, however, is a facultative, within-species variety of LHT that posits evolved mechanisms that developmentally calibrate the suite of phenotypic traits mentioned above, based on environmental cues experienced in childhood. (Nettle & Frankenhuis, 2019) call this LHT-P, or LHT as applied in psychology, and distinguish it from LHT-E, or LHT as used in evolutionary biology.

There has been some criticism of this “facultative” version of LHT (Frankenhuis & Nettle, 2020; Zietsch & Sidari, 2019). One is that it improperly imports theory developed to explain taxonomic differences to explain within-species differences (Stearns & Rodrigues, 2020). These concerns add to the epistemic risk involved in explaining individual differences using LHT. Moreover, the use of LHT to explain within-species differences in humans has racist origins in the use of LHT to explain, for example, putative racial differences in IQ. This, too, raises serious alarms not just about epistemic risk but the downstream ethical risks of using LHT to explain individual differences in psychology and behavior (Sear, 2020).

In addition to these, I would like to consider yet another epistemic risk that in turn entails epistemic risk. This is the possibility that the very same patterns of behavior that LHT theorists explain via evolved facultative mechanisms might have a different explanation proximately, and therefore ultimately. What if (some) people growing up in environments that LHT theorists have described as possessing “cues to short life expectancy” do exhibit some of the psychological and behavioral patterns described, such as greater risk-taking, earlier reproduction, and increased disregard for the future—and yet these are *not* the result of an evolved, facultative calibration mechanism? If we call the evolutionary hypothesis the facultative life history calibration (FLHC) hypothesis, we can call varieties that invoke different proximate mechanisms to explain the individual differences in question “not-FLHC” hypotheses, or ~FLHC. How much does existing evidence uniquely support the FLHC hypothesis, as opposed to varieties of ~FLHC? And are there ethical risks associated with giving excess epistemic credulity to FLHC as opposed to ~FLHC?

What I have in mind, here, is the use of FLHC to explain patterns of behavior associated with poverty and economic and social marginalization and patterns of behavior associated with wealth and social privilege—distinctions that, in places like the U.S., are associated with class and race. Such patterns might include things like higher rates of teenage pregnancy, drug use, crime, gang membership, dropping out of school, and disregard for the future among people growing up in places with poor economic prospects and low life expectancy. Wealthy people, on the other hand—in the U.S. at least—tend to have children later, invest in education, and make plans for secure retirement. LHT theorists have pointed to such patterns as support for the theory, and even conduct studies in which pictures of poor and wealthy neighborhoods are used to nudge peoples' psychological profiles towards "fast" or "slow" patterns in an experimental context (Sng et al., 2017). Importantly, FLHC—the existence of an evolved, facultative calibration mechanism—is the explanation offered for them.

But is this the best, or most likely, explanation? Here, we might want to think carefully about what the alternatives might be—all the possible varieties of ~FLHC. I won't attempt to develop all the possibilities here, but ask yourself whether any of the patterns associated with poverty mentioned above, such as higher rates of teen pregnancy or drug use, could be explained by something *other* than an evolved, facultative mechanism that calibrates these aspects of peoples' psychology as a response to environmental cues of low life expectancy. If you can envision mechanisms—e.g., social, cultural, or historical ones, or ones based on learning via some other mechanism—then you have thought of a plausible explanation that falls within the domain of ~FLHC. For example, the low availability of jobs might lead some to engage in theft or drug dealing because of personal failure to secure a job, or the social proliferation of these options as easily accessible income possibilities, or both. Low incomes might lead one to worry less about saving for retirement, simply because there is no point to it. In contrast, those born into wealth and security might be more likely to seek higher education because they can, or because of culturally transmitted expectations. Individual learning based on personal experience, social transmission of expectations and norms, and the shaping of local options by the historical factors that produce low life expectancy, could all be explanations for patterns that LHT theorists would like to explain by recourse to specialized systems of facultative calibration that are products of natural selection.[4]

Unfortunately, many “tests” of LHT are confirmatory: they ask whether the patterns predicted by LHT exist or not, without considering the plausibility of other mechanisms or processes to explain the same results. Against this backdrop, if you are inclined to think that ultimate-level explanations are inherently “better” than proximate ones, then FLHC becomes an appealing explanation. It might even seem parsimonious, if all of these patterns—crime, drug use, early pregnancy, dropping out of school, disregard for the future—can be “explained” by the same mechanism. In fact, there are purely epistemic reasons to give ~FLHC hypotheses more weight, *a priori,* than FLHC ones, thereby raising the bar for evidence needed to lean in favor of FLHC. One is the large number of alternative proximate explanations for group differences caused by differences in wealth, class, and privilege. Another is the fact that most FLHC studies provide little or no evidence for the actual proximate mechanism proposed. The vast majority simply point to correlations between individual traits and developmental circumstances. Other studies, such as experimental “priming” studies mentioned above, are dubious in that FLHC hypotheses postulate facultative developmental mechanisms, not changes in cognitive patterns on the timescale of minutes after exposure to stimuli such as photographs. Here, the allure of evolutionary explanations seems to be leading researchers to take patterns of evidence as confirming their preferred explanation, without considering what other explanations might be—or, at the very least, not given them a fair empirical shake.

Some might take issue with the proposal of bias towards evolutionary explanations in the LHT case. I accept that I might well be wrong here. However, independent of the epistemic risks, what are the ethical risks to asserting or believing, for example, that patterns of crime associated with poverty are caused by an evolved, facultative developmental mechanism? Are there any, and how serious might they be?

---

[4] In response to this suggestion, I have sometimes heard evolutionary social scientists argue that these other possible proximate explanations are, in the end, the “same” as LHT (FLHC) explanations. I don’t think this is the case. Evolutionary explanations hinge on mechanisms, and for LHT explanations to be true, the observed patterns must ultimately be the result of natural selection specifically to produce those patterns.

Some of the considerations, here, are similar to those for the ADHD case. There is a large literature on how framing the causes of a particular phenomenon changes how people think about it, including what kinds of interventions or remedies they might think are appropriate, or not, for changing it (Jou et al., 1996; Kvaale et al., 2013; Waldmann & Hagmayer, 2013; Zeki et al., 2004). As the ADHD example shows, "biologizing" or "essentializing" a phenomenon can lead to a certain kind of fatalism: if ADHD is "hard-wired," then maybe we can't ever hope to change ADHD itself, only to provide after-the-fact remedies such as medications. Similarly, if crime is a *biological* response to poverty, then perhaps we can never hope to sever the poverty-crime link. Our only options might be either to alleviate poverty, or to be less harsh to criminals. These are both good pathways to social good, in my view. But seeing them as the only options takes for granted that cues to short lifespan will result in crime—or at least, greater risk-taking, shorter time horizons, and so on. What if this isn't true? Our options for intervention might be different. People might think differently about the behavioral patterns and outcomes associated with poverty. And if social interventions can produce harm or prevent it, then being wrong about the poverty-crime link does indeed expose us to ethical risk.

I suspect that this is only the tip of the iceberg of the ethical risks to which we expose ourselves by accepting too quickly LHT explanations for behavior. There are theories of justice and incarceration that hinge on causal theories of behavior. Politicians are elected based on their claims about the causes and consequences of crime and poverty. Economic theories hinge on theories of how and why people choose to do what they do, such as spend or save their money. Not just theories but actual policies, public opinions, and real-world outcomes are influenced by how these matters are framed. It is thus hard to imagine that something as important as explaining phenomena such as crime, reproductive decisions, risk-taking, and investment in one's own life and social capital could not engender ethical risks.

13.7Conclusion: Incentive structures as amplifiers of risk

I have delved deeply into the epistemic and ethical risks associated with evolutionary explanations in the cases of ADHD and LHT, but these are by no means the only ones. There are many, many kinds of evolutionary explanation for human behavior that postulate diverse ultimate causes and proximate mechanisms. We might expect each of these, in turn, to entail their own epistemic and ethical risks. Evolutionary theories of sex differences are an obvious example whose risks have been examined and discussed extensively (Eagly & Wood, 1999). Popular theories of sexual selection such as parental investment theory, for example, have been criticized for offering too narrow a view of the evolutionary functions of sex and gender, leading to the normalization of certain kinds of gender roles and stigmatization of others (Roughgarden, 2013). Beyond ADHD, similar concerns have been raised about biological explanations for phenomena such as autism and schizophrenia, and their consequences for medicalization, interventions, and social stigma (Kvaale et al., 2013; Silberman, 2015). Nor are these concerns limited to "genetic" evolutionary theories of human behavior. They apply to cultural evolutionary theories as well, which posit histories of cultural selection to explain contemporary patterns of behavior, including group differences (Clancy & Davis, 2019; Fuentes, 2022). The answer is not necessarily to avoid thinking about these phenomena through a biological or evolutionary lens. Rather, it is to recognize the possibilities of epistemic and ethical risk, and to take care to broaden our consideration of possible explanations accordingly (Barrett, 2020).

Unfortunately, many aspects of institutional and social life act as amplifiers of the ethical risks that flow from epistemic risk. The seductive allure of evolutionary explanations makes them particularly appealing on an individual level, which makes them particularly effective as viral ideas within and outside academia (O'Connor & Weatherall, 2019). This not only amplifies their spread, but the likelihood that they will be entertained as possible explanations, which in turn increases the chance that they will be *accepted* as explanations, among scientists and the general public alike. Added to this, there are many aspects of the incentive structure of science, and the academy more generally, that "select for," in cultural evolutionary terms, explanations that are easy to consume and appear persuasive and authoritative (Smaldino & McElreath, 2016). Hitting on a seductive but wrong explanation can make you more famous, get you more grant money, and earn you tenure faster than one that is obvious, uninteresting or perhaps too complex for a nice sound bite, but that it is in fact correct.

These are all amplifiers of risk, both epistemic and ethical. Let us be clear: evolutionary explanations are not the *only* ones that entail such risks, and perhaps they are not even the riskiest ones (economic theories, for example, could be even more dangerous). Still, it is perilous for evolutionary researchers to ignore or dismiss the observation that evolutionary narratives carry a particular kind of weight, both in science and popular culture—especially as evolutionary narratives replace religious ones as the origin stories of humankind (Landau, 1993). None of this should stop evolutionary research, nor prevent the proposal and consideration of evolutionary hypotheses. It should, however—as in all science and all scholarship--make us think carefully about what we are doing.

Acknowledgments

Thanks to André Ariew, Karthik Panchanathan, Daniel Nettle, Thom Scott-Phillips, and other participants at the GenDar workshop on Evolutionary Thinking Across Disciplines in Paris in 2021 for invaluable insights and commentary on ideas presented here.

References

Babic, B. (2019). A theory of epistemic risk. *Philosophy of Science*, *86*(3), 522–550.

Barrett, H. C. (2015). *The shape of thought: How mental adaptations evolve*. Oxford University Press.

Barrett, H. C. (2020). Deciding what to observe: Thoughts for a post-WEIRD generation. *Evolution and Human Behavior*. https://doi.org/10.1016/j.evolhumbehav.2020.05.006

Bernhardi, R. von, Bernhardi, L. E., & Eugenín, J. (2017). What is neural plasticity? *The Plastic Brain*, 1–15.

Chang, F.-M., Kidd, J. R., Livak, K. J., Pakstis, A. J., & Kidd, K. K. (1996). The world-wide distribution of allele frequencies at the human dopamine D4 receptor locus. *Human Genetics*, *98*(1), 91–101.

Clancy, K. B., & Davis, J. L. (2019). Soylent is people, and WEIRD is white: Biological anthropology, whiteness, and the limits of the WEIRD. *Annual Review of Anthropology*, *48*, 169–186.

Copping, L. T., Campbell, A., & Muncer, S. (2013). Impulsivity, sensation seeking and reproductive behaviour: A life history perspective. *Personality and Individual Differences*, *54*(8), 908–912.

Currie, A. (2018). *Rock, bone, and ruin: An optimist's guide to the historical sciences*. MIT Press.

Curry, O. (2006). Who's afraid of the naturalistic fallacy? *Evolutionary Psychology*, *4*(1), 147470490600400130.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559. https://doi.org/10.1073/pnas.1517441113

Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science*, *67*(4), 559–579.

Douglas, H. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Pre.

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, *54*(6), 408.

Eisenberg, D. T., Campbell, B., Gray, P. B., & Sorenson, M. D. (2008). Dopamine receptor genetic polymorphisms and body composition in undernourished pastoralists: An exploration of nutrition indices among nomadic and recently settled Ariaal men of northern Kenya. *BMC Evolutionary Biology*, *8*(1), 1–12.

Elman, J. L., Bates, E. A., & Johnson, M. H. (1996). *Rethinking innateness: A connectionist perspective on development* (Vol. 10). MIT press.

Frankena, W. K. (1939). The naturalistic fallacy. *Mind*, *48*(192), 464–477.

Frankenhuis, W. E., & Nettle, D. (2020). Current debates in human life history research. *Evolution and Human Behavior*, *41*, 469–473.

Friedman, R. A. (2014, October 31). A Natural Fix for ADHD. *The New York Times*. https://www.nytimes.com/2014/11/02/opinion/sunday/a-natural-fix-for-adhd.html

Fuentes, A. (2022). WEIRD Indeed, but there is more to the story: Anthropological reflections on Henrich's "The Weirdest people in the world." *Religion, Brain & Behavior*, 1–58.

Glavin, G. B., & Szabo, S. (1990). Dopamine in gastrointestinal disease. *Digestive Diseases and Sciences*, *35*(9), 1153–1161.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *205*(1161), 581–598.

Hempel, C. G. (1954). A logical appraisal of operationism. *The Scientific Monthly*, *79*(4), 215–220.

Jou, J., Shanteau, J., & Harris, R. J. (1996). An information processing view of framing effects: The role of causal schemas in decision making. *Memory & Cognition*, *24*(1), 1–15.

Kvaale, E. P., Haslam, N., & Gottdiener, W. H. (2013). The 'side effects' of medicalization: A meta-analytic review of how biogenetic explanations affect stigma. *Clinical Psychology Review*, *33*(6), 782–794.

Landau, M. (1993). *Narratives of human evolution*. Yale University Press.

Lewontin, R. (1996). *Biology as ideology: The doctrine of DNA*. House of Anansi.

Moore, G. E. (1959). *Principia ethica:(1903)* (Vol. 2). Cambridge University Press Cambridge.

Motta, M., Callaghan, T., & Sylvester, S. (2018). Knowing less but presuming more: Dunning-Kruger effects and the endorsement of anti-vaccine policy attitudes. *Social Science & Medicine*, *211*, 274–281.

Nettle, D., & Frankenhuis, W. E. (2019). The evolution of life-history theory: A bibliometric analysis of an interdisciplinary research area. *Proceedings of the Royal Society B*, *286*(1899), 20190040.

O'Connor, C., & Weatherall, J. O. (2019). *The misinformation age: How false beliefs spread*. Yale University Press.

Pritchard, D. (2016). Epistemic risk. *The Journal of Philosophy*, *113*(11), 550–571.

Rafalovich, A. (2013). Attention deficit-hyperactivity disorder as the medicalization of childhood: Challenges from and for sociology. *Sociology Compass*, *7*(5), 343–354.

Roughgarden, J. (2013). *Evolution's rainbow: Diversity, gender, and sexuality in nature and people*. Univ of California Press.

Sear, R. (2020). Do human 'life history strategies' exist? *Evolution and Human Behavior*, *41*(6), 513–526. https://doi.org/10.1016/j.evolhumbehav.2020.09.004

Silberman, S. (2015). *Neurotribes: The legacy of autism and the future of neurodiversity*. Penguin.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*(9), 160384.

Sng, O., Neuberg, S. L., Varnum, M. E. W., & Kenrick, D. T. (2017). The crowded life is a slow life: Population density and life history strategy. *Journal of Personality and Social Psychology*, *112*(5), 736–754. https://doi.org/10.1037/pspi0000086

Stearns, S. C., & Rodrigues, A. M. (2020). On the use of "life history theory" in evolutionary psychology. *Evolution and Human Behavior*.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift Für Tierpsychologie*, *20*(4), 410–433.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293.

Waldmann, M. R., & Hagmayer, Y. (2013). *Causal reasoning.*

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*, *20*(3), 470–477.

Zeki, S., Goodenough, O. R., Fugelsang, J. A., & Dunbar, K. N. (2004). A cognitive neuroscience framework for understanding causal reasoning and the law. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1451), 1749–1754.

Zietsch, B. P., & Sidari, M. J. (2019). A critique of life history approaches to human trait covariation. *Evolution and Human Behavior*.

Chapter 14

Evolution and Ecology of Organizations and Markets

Randall E. Westgren[i]

Abstract

I examine the prospect of locating the evolution of economic organizations and markets within Generalized Darwinism. I use the most mature account Joseph Schumpeter's model of economic evolution form his 1939 book, *Business Cycles*, as the basis for the analysis. The benefits of this account are that it abstracts variation as innovation -- purposeful entrepreneurial action to create novel technological and organization processes and structures – and idealizes the variation as a new production function, a construct central to economic analysis. New firms and reconfigured existing firms are equivalent, an advantage for theorizing. Measured against these benefits for evolutionary thinking at the population level is the complex combination of competition and selection processes within a community ecology of economic mutualisms, conventions, institutions, and other inter-firm structures that confound the explanation of the success of the variants. Furthermore, if one takes the purposeful innovation as the organizational strategy, modern accounts of competitive advantage as the measure of fitness require opacity of the strategy to other organizations. This renders typical accounts of diffusion common to cultural evolution moot. Thus, the selection and retention features of current accounts of Generalized Darwinism are incompatible with organizational evolution.

Keywords : Economic evolution, Schumpeterian evolution, organizational ecology, resource-based view of the firm, innovation.

---

[i] Division of Applied Social Sciences, University of Missouri-Columbia, USA, Division of Applied Social Sciences. Email: westgrenr@umsystem.edu

The study of organizations and populations of organizations in economics and sociology has a long history. During the last third of the 19th Century and the first third of the 20th Century, social scientists actively carried out analyses of the dynamics of social change while actively spanning the putative boundaries of economics, sociology, and political science (Scott 2014). The institutionalization of social rules, structures, organizational forms, and artifacts and the de-institutionalization of others was clearly an evolutionary process, broadly defined. During the 20th Century, the boundaries between these social sciences became institutionally defined even though they shared some common scientific paradigms including more individualistic approaches and common exploitation of rational actor theory. Scott (2014) identified these shared approaches as neo-institutional theory. Historical methods declined, especially in economics, which increasingly focused on highly idealized models of individual consumer choice and highly idealized models of profit-maximizing firms, which meet in utterly abstract markets wherein maximizing behaviors yield equilibria.

During the second half of the last century, the study of organizations, including firms, became the province of schools of business and this scholarship drew on social psychology, sociology, and (heterodox) economics. The business school scholars were often trained in one of the disciplines but operated in a transdisciplinary environment. Organizational studies of firm life cycles, inter-firm strategies (competition, but also mutualisms like joint ventures and cooperatives), and the social construction of industry and product categories had temporal dynamic features. Because the units of analysis varied across individuals, work teams, firms, strategic groups (direct rivals), industries and sectors; and because the levels of analysis varied from individual (person, team, firm, etc.) to collections/populations of units we find that the concept of evolution was invoked across all the units and levels of analysis. Firms evolve. Industries evolve. Markets evolve.

Importantly, inter-firm behaviors also seem to evolve. Many parts of modern economies have seen traditional physical markets with visible exchanges displaced by electronic markets between anonymous agents. One-off "open" market transactions are replaced by contracts, often very long-lived. The humorous meme of being able to purchase groceries on-line while not wearing pants became a behavioral norm during the COVID-19 pandemic when shops, restaurants, and farmers markets were closed. New organizations compete actively for their share of deliveries for a growing market for on-line restaurant purchases delivered to the buyer's door. Simple markets with traditional vendor-buyer exchanges have become more complex. New forms of mutualisms arise, increasing the complexity of studying a sector of the economy.

If we are to assess the possibility that the study of the dynamics of organizations and markets can fit comfortably under the umbrella of Generalized Darwinism, it is necessary to parse out the issues of which elements of the economic system are evolving in a manner consistent with other social sciences, as well as biology. To begin the task in this paper, I first take a book by G. Evelyn Hutchinson, *The Ecological Theater and the Evolutionary Play*, as a touchstone for my argument. Hutchinson was a force for the mathematization of ecology and the consideration of the information flows across the boundary between ecology and evolution. One sees the latter phenomenon under the rubrics of biogeography, niche construction, and evolutionary ecology. In this paper I return periodically to the need to consider horizontal (ecological) connections among the organizations in an economic sector as one considers vertical (evolutionary) temporal processes. In the next section of the paper, I give a necessarily brief review of evolutionary thinking about organizations, notably firms, in economics and management in the 20th Century. The lack of a consensus about how to approach evolutionary thinking about organizations and markets causes me to return to the best-known scholar in evolutionary economics, Joseph Schumpeter, in a re-examination of his account of innovation as the foundation for economic evolution. I exploit Schumpeter's abstraction to sort out the nature of evolutionary processes at the analytical level of firm populations which extend to units of analysis including strategic groups and inter-firm behaviors. To conclude, I consider the degree to which these dynamic phenomena can be accommodated in Generalized Darwinism.

14.1Brief History of Evolutionary Thought in Economics and Management

Thorstein Veblen wrote a polemical piece for the *Journal of Political Economy* in 1898, "Why is Economics not an Evolutionary Science?" Veblen was a young professor of economics at the young University of Chicago; he was a member of the founding group in the social sciences, following the lead of J. Laurence Laughlin, who was advising Veblen's graduate degree in economics at Cornell University. Veblen never completed a degree in economics. He held a PhD in Philosophy from Yale, where he studied under the evolutionary sociologist William Sumner in the tradition of Herbert Spencer. Veblen was given the editorship of the *JPE*, which allowed him a forum for his ideas at the fuzzy boundary between sociology and institutional economics. In the 1898 article, he railed against classical economics, Austrian Economics, and the German Historical School as static science and made a rather incomplete case for a taxonomic approach (I presume a phyletic approach) to unfolding changes in social (economic) phenomena. He continued this line of thought for the remainder of his career. His most important works were more sociological and anthropological than economic and he is counted among the most radical of heterodox economists in the history of economic thought.

We use Veblen's polemic as a touchstone for the history and philosophy of economics. Two decades into the 21st Century, we may repeat his question without a satisfactory answer. Not much work was done on evolutionary economics for first three decades of the 20th Century as the science of economics passed from classical to neoclassical doctrine that celebrated idealized models of consumer behavior and firm behavior in static equilibrium. To this day, this is orthodox economics for both "high science" and instruction of undergraduates. Economics passed from verbal accounts of market processes in Veblen's time to mathematical accounts of highly idealized markets by the 1940s[1]. Historical accounts of market processes gave rise to institutional economics in this same period, but this became heterodox economic science with diminished professional status.

Evolutionary economics was resuscitated very briefly by Kenneth Boulding in two books, *A Reconstruction of Economics* (1950) and *Ecodynamics: A New Theory of Societal Evolution* (1978). His approach took a more ecological stance than an evolutionary one and relied on vivid analogies with biology. Moreover, he insisted that the analysis of human systems must be

---

[1] The milestone in this history is the publication of Paul Samuelson's *Foundations of Economic Analysis* (1947). It was the clarion call for formal mathematical models which became the *sine qua non* for training economics graduate students and for respected publications in the profession. One of the few critics of this phenomenon was Kenneth Boulding.

founded on both biology and economics. He was a stalwart in the burgeoning societal awareness of environmental issues in the 1970s and 1980s, but his work had no measurable impact on the economics profession.

Following the publication of Boulding's 1950 book, there was a brief and vigorous debate on the value of biological analogies in economics. Two serious antagonists, Armen Alchian of UCLA and Edith Penrose of Johns Hopkins, wrote back and forth in the flagship (orthodox) journal of the American Economics Association in 1952 and 1953.[2] The debate was not carried on by economists until an excellent article by Jack Hirshleifer was published in 1977 as a response to E.O. Wilson's *Sociobiology: The New Synthesis* (1975). Hirshleifer considers economics to be a part of the sociobiology realm along with other social sciences and that "The fundamental organizing concepts of the dominant analytical structures employed in economics and in sociobiology are strikingly parallel... Fundamental concepts like scarcity, competition, equilibrium, and specialization play similar roles in both spheres of inquiry. And terminological pairs such as species/industry, mutation/innovation, evolution/progress, mutualism/exchange have more or less analogous denotations." (Hirshleifer 1977, p. 2) He goes on to examine these isomorphic constructs in detail while considering possible breakdowns around fitness and maximization.

It is important to recognize that the debates of the middle of the 20th century were not focused entirely on evolutionary biology as a source of analogical reasoning in economics. Most of the economists reaching across the boundary to natural science included the elements of ecology such as mutualisms and competition, growth, and intra-specific relationships inherent in sociobiology. It was during the 1980s that evolutionary economics regained the interest of heterodox economists with new book series, new journals in institutional and evolutionary economics, and appreciation for firm and market dynamics in business schools,

---

[2] I recommend Clement Levallois (2011) on this debate. Levallois also places the debate in the context of Generalized Darwinism.

where scholars drew from economics, sociology, and social psychology. A sociological approach to firm population dynamics was particularly evident during the 1980s and 1990s, labeled population ecology of organizations (cf. Hannan and Freeman 1977, 1989; Carroll 1988; Delacroix, Swaminathan, and Solt 1989; Swaminathan and Delacroix 1991). This literature focused on long time series of organizational population dynamics of a single organizational form without speciation and limited inter-population competition or mutualism. A retrospective view of this literature shows limited value in the study of evolutionary processes despite some disagreement from supporters of Generalized Darwinism (Reydon and Scholz 2009, Scholz and Reydon 2008).

On the other hand, economic models of industry dynamics -- typically labeled as evolutionary economics -- were plentiful. It is beyond the scope of this paper to complete a review of this burgeoning literature. I rely on a careful analysis of the ontological, heuristic, and methodological issues that defined this literature by Ulrich Witt (2008). Witt creates a 2 x 2 matrix that is useful in characterizing the competing approaches. See Figure 1 below.

***Table 1: Classifying the Evolutionary Economics Literature**. Source: Witt (2008) Figure 1 pg. 555.*

HEURISTIC STRATEGY	ONTOLOGICAL STANCE	
	MONISTIC	DUALISTIC
Generalized Darwinian Concepts (variation, selection, retention)	UNIVERSAL DARWINISM	NEO-SCHUMPETERIAN
Generic Evolutionary Concept (novelty emergence, dissemination)	NATURALISTIC APPROACHES	SCHUMPETER (1911)

There are three quadrants in Witt's matrix that I consider in this paper. The monistic approach to a common form of evolutionary theory, based upon variation, selection, and retention is labeled as Universal Darwinism, now more universally called Generalized Darwinism. This project seeks an approach to evolutionary social science that has some common features to the Darwinian account in biology. This volume is dedicated to understanding the nature of this project. I return to this quadrant later in this paper. A dualistic approach that uses the Darwinian concepts within the framework of economics is a large literature that Witt and others label as Neo-Schumpeterian. There is distinction between the nomenclature and the processes in these economic accounts and the biological accounts, though there are metaphorical devices that link them. But the intent of the economics literature to examine the phenomena of evolution in industries within the broader field of economic science and not as evidence of universality across the natural and social sciences. In this literature it is the norm to find mathematical models, often as computer simulations through time steps, that track industrial dynamics, competition, firm and industry growth and other outcomes from innovation and technology adoption. These accounts are Neo-Schumpeterian in the sense that the original account of industrial evolution rests incontrovertibly with Joseph Schumpeter, an Austro-American economist who first wrote of this in 1911 in *Theorie der wirtschaftlichen Entwicklung* and throughout the remainder of his career.

For my purposes, I must describe Schumpeter's model to further consider the classification boundary the Witt invokes to place it at the opposite corner from Generalized Darwinism. This is necessary because Witt bases his classification on the 1911 Schumpeter and a book Schumpeter published in 1942: *Capitalism, Socialism, and Democracy*. The latter book is Schumpeter's most-read work. It is a very broad treatise in social science, wrapping his model of economic evolution inside of social and political theory. The author considered the work a "potboiler" and not a serious work in economics. In the section which follows, I clarify the Schumpeterian model of economic evolution.

14.2 Schumpeter's Evolutionary Economics[3]

---

[3] I have unabashedly stolen this heading from the title of Esben Sloth Andersen's monumental book (2009) on the evolution of Schumpeter's models from 1908 to 1950.

Joseph A. Schumpeter wrote several books to fill the need for a theory of economic progress that could not be accounted for in the prevailing classical and neoclassical theories of economics that were, and remain, essentially static. The prevailing theories were focused on the economic processes that distributed value (i.e. the market price for a good) to the factors of production: land, labor, and capital. Schumpeter saw this as a stationary system, with modest adjustments for exogenous changes (natural phenomena such as drought, population changes, war, and other governmental action), taken as responses to data without changing production technologies, market structures, or consumer behavior. The modest adjustments permitted the system to return to stasis that approximated the static state prior to exogenous shocks.

Schumpeter needed a mechanism for an economy (perhaps an industry or sector) to evolve to a higher economic state, measured by gross domestic product (GDP), per capita incomes, or a similar measure. One would naturally observe higher incomes for laborers, increased consumption of goods by households, or wealth accumulation in the higher economic state. These things cannot be observed in the circular flow of a market, industry, sector, or nation in equilibrium. In his 1911 book, *Theorie der wirtschaftlichen Entwicklung*, the mechanism he chose was the action of the entrepreneur who used new combinations (*neue Gestalten*) to break away from the static market. The entrepreneur earns an entrepreneurial profit (*Unternehmergewinn*): the difference between the costs of creating the new combination and its value in the market. This account by a young scholar was very much of a Nietzschean "man of action" overcoming resistance by the masses.

The 1911 book has never been translated fully into English. A second edition in German (1926) was revised radically, favoring entrepreneurship as an economic function rather than a characteristic of individuals. The second edition was the basis for a translation as *The Theory of Economic Development* (1934) after Schumpeter's arrival in the United States. This book is widely read as the expression of Schumpeter's account of innovation as the disequilibrating mechanism that moved the economy from a lower level of stasis to a higher level (of GDP, for example). The term "development" was chosen to avoid problems associated with social Darwinism at the time; the German term *Entwicklung* translates equally as development and evolution.

For the remainder of this section of the paper, I rely upon the last of Schumpeter's books that focus on evolutionary economics, *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process* (1939). This book contains the clearest and most elaborate presentation of his distinction between the static model of stationary flow and (as Chapter 3 is titled) How the Economic System Generates Evolution. Notably, Schumpeter finally declares his account to be about evolution and not development. He clarifies the mechanism of innovation to be several forms of "the setting up of a new production function". The production function has been a mainstay of economic modeling since the late 1900s. It permits economists to move easily between verbal, graphical, and mathematical accounts of the technology (widely defined) that use production inputs or factors to create products or outputs. This is best demonstrated by one classic functional form, The Cobb-Douglas production function:

$Y = AK^aL^b$, where

Y is the number of units of the product,

K is the number of units of capital goods (however measured),

L is the number of units of labor (however measured),

a and b are coefficients that relate the proportions of K and L and represent input/output relations

between capital and labor to product, respectively, and

A is total factor productivity – an efficiency measure.

The "technology" described in the Cobb-Douglas function is a combination of A, a and b. The coefficients a and b create the interesting result of returns to scale. That is, if a and b sum to 1, the process has constant returns to scale; a doubling of the two inputs results in a doubling of output Y. If $a + b < 1$, then output rises less than proportionally (decreasing returns to scale). And if $a + b > 1$, then output rises faster than inputs. Finally, one sees that the coefficient of efficiency A also affects the overall input-output relation. Any innovation in the production process that raises a, b, or A relative to existing production technology creates the opportunity for the innovator to earn entrepreneurial profits. If the innovation is significantly large, the innovator may sweep away less efficient incumbent firms in a result Schumpeter calls "creative destruction" (schöpferischen Zerstörung).

The isolation of the mechanism to be wholly about the production function admits to some broader conceptions of production and innovation than the foregoing. Schumpeter, in the 1939 book, allows for innovations to be the creation of a new good (typically from different inputs), a new form of organization (e.g. on-line vs. brick-and-mortar retail, outsourcing components in less-developed economies, or conglomerate mergers), or perhaps producing an existing good in a new production function such as organic, low carbon footprint, or other sustainable method. Thus, within the isolation of innovation as a change in the production function, there are several "how possibly" forms of the mechanism. In Ross and Westgren (2006), there are several innovations that are not of the technical coefficients of the production function *per se* but include structural changes in organization design (lowering transaction costs), and in organizational governance, and product innovation (new attributes). In any case, the innovation results in superior performance of the innovating firm vis-à-vis incumbent firms. Additionally, a modern conception of the firm-as-production-function includes multiple kinds of capital including intangible assets such as intellectual property and brand names; and multiple types of human resources (e.g. R&D teams, logistics experts, etc.). Finally, there may be technology variables that directly affect product attributes and the nature of inter-firm competition occurs not with homogeneous products, but differentiated, heterogeneous goods (Westgren and Wuebker 2019).

Schumpeter makes an important idealization in his verbal model, particularly in the 1939 book. He recognizes that some incumbent firms may innovate by changing their production function, but he counts these as New Firms. This idealization may not sit well with scholars who follow case histories of often long-lived firms. But the benefits of this idealization include the following. (1) We no longer must resort to Lamarckian accounts of long-lived economic organisms that acquire and retain new production attributes. (2) With population thinking, we are not concerned with identification of individual firms, just the distribution of attributes (e.g.

production technology) that affect population fitness. (3) This allows us to deal with very large firms that evolve, which is a difficulty for population-level evolutionary theory when the population is singular. As an example, consider Amazon. Clearly, Amazon of 2022 is not the same organization as Amazon of 1995 – an on-line seller of books. Since 1998, this organization has made 108 acquisitions of other firms and 22 joint ventures (mutualistic structures). This is much more than organismal plasticity. But if we conceive of Amazon as a collection of many production functions, we may disaggregate the legal structure and observe Schumpeterian processes as these parts compete as New Firms in multiple markets.

There are two additional features of the 1939 Schumpeter account that contribute to a more complete evolutionary economics. First, the process is more than the variation among competing organisms (production functions). The New Combinations are introduced into the Economic Organism, Schumpeter's term for the industrial and social ecology surrounding the innovations (embodied as New Firms). New products (or known product classes with new attributes) are assessed by buyers, who make a subjective assessment of the products in comparison with existing products and their prices. Some of these will not be selected for by buyers and fail in the market because they are insufficiently differentiated or have a price disadvantage. This translates into failure of the New Firm, who is unsuccessful in harvesting resources (product revenues) from the Economic Organism. Other New Combinations will fail to obtain another important resource from the Economic Organism: financial capital to sustain the new enterprise. And one sees the common form of competition -- apparent competition by two organisms that thrive on a common resource -- causes differential success in harvesting resources (revenues or capital) because of the differences between their production functions. The more efficient New Firm will survive in repeated competition; the other will die.

The second feature of this account is that innovations tend to occur in swarms, both within a given industry and in adjacent industries. An innovation that is sufficiently visible to competitors will engender replication, though with error. Some elements of any New Combination will be opaque to potential adopters and some elements of the innovation -- especially not tied to physical assets -- will not be replicable. I return to this issue later in the paper. On the other hand, another New Firm may copy elements of the first innovation and make improvements in design or efficiency. It may surpass the initial innovator within a few time periods. Where a swarm of innovations opens up a new market, the potential payoffs will attract many experimental New Firms and inter-firm competition within this unsettled landscape of the Economic Organism will be rivalrous. Schumpeter argues that in these cases, adjacent industries in the Economic Organism will seek to adapt the innovation to their industry. The development of GPS tracking for transportation (shipping) firms led to the adaptation of GPS for railroads and long-distance trucking. As well, firms in vertically adjacent industries such as suppliers to the New Firms will adapt in a form of derived innovation (Redlich 1951). Workers (suppliers of labor) may require new training to fit into the New Combinations. Machinery suppliers may have to innovate their production functions to produce new specialized physical assets for their client New Firms. Schumpeter saw this as an evolutionary play in the ecological theater over time and space (regional diffusion, inter-industry diffusion, and vertical diffusion in what we now call supply chains -- the economic analogue of trophic levels in ecology).

Returning to the problem of Amazon, we may characterize that the multiple New Firms that comprise that conglomerate are competing in adjacent markets as part of the swarm. The ecological connections between these related New Firms may create the same kind of linked coevolutionary processes that we observe in natural ecosystems. These coevolutionary processes are common to market ecologies and exploiting the Schumpeterian idealization of firms as production functions may allow us to consider the dynamics within large firms like Amazon and Apple as linked evolutionary processes among smaller units of analysis.

The evolutionary play is not scripted. Access to financial capital and buyer acceptance vary widely across industries. Some industries and some firms will fail in their attempts to innovate. Note that Schumpeter always makes a strong distinction between invention and innovation. The latter is an economic function that goes beyond technical change in products or processes and must include market entry. The market, as part of the Economic Organism, is where "the economic score is kept". And the rivalry that occurs between Old Firms and New Firms doesn't have a foregone conclusion. The ecological theater, Schumpeter's Economic Organism, is complex. The play proceeds on multiple stages simultaneously, often in different venues with different marquees. The competition among these performances creates havoc in important ways. Labor from Old Firms is being recruited to New Firms, affecting wage rates. Fast-growing New Firms need capital and their success means another industry or sector is capital-starved. Suppliers, retailers and other associated firms are faced with derivative increases or decreases in demand and cost that are beyond their ability to control.

For Schumpeter, this phase of economic evolution is a period of adjustment to an innovation swarm. Losers are winnowed out from successful survivors of this turmoil. There will be displaced labor and capital that may find a new higher-payoff place in the new Economic Organism that results. The turmoil will recede, and a new quasi-equilibrium will be reached. Labor and capital that have been absorbed into New Firms should be at a higher wage and rent, respectively. Other labor and capital resources will be idle or under-employed. Entrepreneurs will have earned temporary profits (returns to innovation) that are eventually reduced by the diffusion of the New Combinations across competitors and other industries to the point where the novelty no longer supports higher revenues over costs. New Combinations are now Old Combinations. It is important to recognize that Schumpeterian evolutionary economics is not a teleological account. Granted, the innovating firms are doing so purposely with foresight. But the competitive process within the Economic Organism is complex and outcomes cannot be foreseen. Old Firms may prevail in some sectors of the economy because they have advantageous mutualisms with other firms in adjacent industries that prevent New Firms from becoming profitable. The complexity of inter-firm and inter-industry mutualisms may cause financial capital to be withheld from some innovating firms or starve their appetite for physical and human resources. At the end of a Schumpeterian cycle, the sector or economy may not have grown or evolved in a discernable way.

The cycle will be repeated over time. At the population level(s) of industries, sectors, regional economies, and national economies, the resulting stasis will create the opportunity for entrepreneurial action -- innovation -- to break out of the circular flow to create new entrepreneurial profits. This resembles Eldredge and Gould's punctuated equilibrium (1972). In fact, Schumpeter uses much the same argument as Eldredge and Gould; he contrasts his account to Alfred Marshall's economic gradualism (slow, modest changes that entail marginal adjustments to equilibrium) as Eldredge and Gould contrast their evolutionary account to phyletic gradualism. However, it must be emphasized that Schumpeter never accepted biological evolution as a metaphoric or analogic account of economic evolution.

14.3What is Evolutionary about Schumpeterian Economic Evolution?

We return to Witt's (2008) Figure 1. It seems clear that Schumpeter's account, updated by the 1939 model and its more catholic approach to innovation, still fits comfortably in the southeast quadrant given its dualistic ontological stance and a heuristic strategy of novelty emergence and dissemination. Perhaps. My skepticism turns on what constitutes dissemination. The processes described in Hayek, Georgescu-Roegen, Veblen, and North are truly diffusion. And for these writers, the diffusion process was among humans in what is generally cultural evolution -- a horizontal process that does take human-scale time. Witt is correct that this kind of dissemination requires some visibility of the novel meme which can be transferred with sufficient degree of replicability. The Schumpeterian innovation is doubtless a form of novelty, but dissemination is not as blithe as a cultural meme. Firstly, the firm surrounds the innovation, masking the true nature of the innovation in the production function. A recent literature in management takes this as the centerpiece of the nature of inter-firm competition. Labeled the resource-based view of the firm (RBV), firms establish competitive advantage over rivals by having a portfolio of resources (physical, human, and social) different from the rivals that is valuable (profit-generating), rare (idiosyncratic) and inimitable (Barney and Clark 2007). Inimitability is supported by three phenomena: *path dependence, social complexity, and causal ambiguity*. Path dependence means that the social resources (decision processes, reputation, shared tacit knowledge) cannot be easily replicated by competitors who have not trod the same historical path of accumulating or building these processes. Social complexity is another feature of the "isolated" organization; the human-machine interactions, group processes, and internal structures are not easily understood by outsiders. And causal ambiguity is the natural result of the organizational carapace that surrounds the firm: what is the real source of competitive advantage? Is it the physical assets and computer technology? Or is it the social structures?

Schumpeter's admission of non-technical innovation means that the core of the RBV has meaning for understanding that replicating innovations between firms is fraught. At best, copies are incomplete or misconstrued. Copies may be truer to innovations in small firms, simple firms. This is borne out by a well-understood phenomenon in production agriculture. Willard Cochrane identified this phenomenon in post-WWII American agriculture (Cochrane 1958). Small farms with simple production functions were the population of relevant firms, especially in cereal production. Entry and exit of farmers into the market was easy -- a classic structure of neoclassical economics in the static equilibrium. Prices were always low, sometimes above and sometimes below the cost of production -- profits were often negative for many individuals in the population. Farmers sought to reduce production costs, so as to seek profits. New (larger) machinery, new hybrid seeds, irrigation, and other inventions were readily available for adoption. Farmers that adopted were Schumpeterian innovators; the farm production function was changed to create lower costs. Alas, the inventions were widely available and neighboring farmers could readily observe yield changes, planting and harvesting efficiencies, and other artifacts of the innovation. Replication of the innovation (adoption) followed quickly. The population-level outcome was tragic. Efficiencies created increased supplies of the cereals faster than they could be absorbed by buyers. Prices dropped below previous equilibrium levels -- down to the new level of costs. Profits disappeared. This provoked the next cycle of innovation and lower prices. Cochrane dubbed this phenomenon the Agricultural Technology Treadmill. The ease of replication of New Firms prevented anything that looked like economic progress.

The Cochrane treadmill is a vivid example of diffusion. Forgive *le jeu de mots*. The fact that the innovation was so vivid to potential copiers makes the lesson vivid. In my view, this is an ecological phenomenon where any competitive advantage is fleeting. It has no real impact on the fitness of the organization or, at the population level, the organizational form. There were 6 million farms in 1945. There were 2.02 million identified in the 2020 Census of Agriculture (USDA 2020).

For the Schumpeterian model to work, diffusion cannot be the mechanism in the evolutionary process. Otherwise, we always get a version of the Technology Treadmill. For Schumpeter, part of the fitness of a New Firm is vested in the opaqueness of the new production function. The logic of the RBV, devised a half-century after Business Cycles, is the mechanism that defines competitive advantage (superior resource portfolio) and is the mechanism that retards replication by competitors (inimitability). These strategic cognitive mechanisms -- the innovation strategy of the firm -- are faced with significant uncertainty of outcomes. The bounded rationality of the firm's human agents cannot predict how competition will look. After all, they cannot perceive what is transpiring inside the protective carapaces of their rivals by the same logic of the RBV. Moreover, the structure of the ecology of the Economic Organism is complex and the embedded mutualisms and trophic connections are impossible to know. The process that occurs between entry of the New Firms and the ecology around them may be a selection process. Perhaps.

If there is a selection process in Schumpeter's account, then it is a candidate to be moving to the northeast quadrant of Witt's matrix. As Hodgson and Knudsen (2010) repeatedly note in chapter 5 of their book, selection is a particularly elusive construct when dealing with industrial dynamics. Their base definition of selection, taken from Price (1995), is as follows.

> *Selection* involves an anterior set of entities that is somehow being transformed into a posterior set, where all members are sufficiently similar to some members of the anterior set, and where the resulting frequencies of posterior entities are correlated positively and causally with their fitness in the environmental context. The transformation from the anterior to the posterior set is caused by the entities' interaction within a particular environment. (Hodgson and Knudsen 2010, p. 92)

The Schumpeterian economic evolution process begins with an anterior set of entities: Old Firms and New Firms. The posterior set contains the same entities in different proportions, though at the limit only one of the two subsets may be evident. Schumpeter's creative destruction can work to eliminate Old Firms in a given population (industry). His favorite swarm was the railroadization of the United States and England. Mail coaches were swept away by the railroads. But it is clear to economic historians that incomplete replacement of Old Firms by New Firms is the norm. The transformation of the entities in the population did occur by their interaction with the particular environment: the Economic Organism. This leaves only three subtleties. Are the anterior and posterior populations sufficiently similar? Yes. The competition for resources requires that they must be substitutable rivals in the eyes of buyers and financiers, or the New Firms have no place in the ecology. Are the frequencies of the posterior entities positively and causally correlated with their fitness in the environmental context? I doubt both correlations can be observed. The multiple selection levels noted by Hodgson and Knudsen imply difficulty in establishing causality and the confounding of inter-firm competition with embedded mutualisms (co-evolution) raise doubt about whether correlation can be presumed as positive. This leads to the third subtlety. What is fitness? It certainly isn't as clear as reproductive success in biology. Is it profitability? Short-term survivability? The simplicity of neoclassical economics that relies on maximizing profits is a useful idealization for equilibrium models, but fails in the complex dynamism of evolutionary processes. Nelson and Winter (1982a) identify the choice between innovativeness and efficiency as the "Schumpeterian tradeoff". Moreover, casual empiricism shows that firms follow a multiplicity of objective functions -- not always as maximizers -- including revenue growth, market share, social missions, and stakeholder (employees, community, etc.) payoffs.

My ambivalence toward selection as a meaningful characteristic of economic evolutionary accounts, including Schumpeter's, rests on the two subtleties discussed above and on another feature of both the Witt (2008) and Hodgson and Knudsen (2010) accounts: *routines*. There is a long debate on the meaning of the term and whether any of the meanings can contribute to evolutionary thinking (Becker 2004; Becker, Lazaric, Nelson & Winter 2005; Parmigiani & Howard-Grenville 2011; Vromen 2006, 2011). The term has been long used in management, especially organization theory, to describe embedded processes. The original use of the term, ascribed to Herbert Simon, is for routinized, repeatable processes that do not require significant judgment or cognitive capacity by the human agents that are involved. Such processes are nearly automatic, as in responding to, "Please pass the salt." Over time, more complex processes have been called routines. Evolutionary theorists from management and heterodox economics have placed these processes interior to organizations subject to evolutionary pressures following Nelson and Winter (1982b). The value of this construct for evolutionary modeling is that it takes on the role of genes and appears as a replicator, a target of selection, a carrier of organizational inertia, and the source of variation. Routines are copied within firms, between firms, and between populations of firms. This frustrates a sound approach to testing the fit between Generalized Darwinism and any model of organization evolution.

If I seek to fit the construct of routines into the Schumpeterian account, there are two strategies. First, routines are the building blocks of the production function and are thus embodied in Schumpeter's model of the firm in the evolutionary processes. Second, as Schumpeter has chosen the production function as the central construct in defining innovation in the firm, routines are superfluous. I am fond of the simplicity of the second strategy, but I will deal with the first strategy. Given that the routine can be highly visible and copiable, it fails the Technology Treadmill test. If the routine is the stable locus of retention in the firm, it cannot also be the source of variability within the firm. If the former, can it serve the purpose of the Schumpeterian innovator? If the latter, can we expect it to be present in both the anterior and posterior entities after the evolutionary process? Finally, in the competitive process among Old and New Firms, are routines sources of competitive advantage -- valuable, rare, and inimitable? No, especially if they are easy to observe and understand and are thus replicable.

I do not feel comfortable that economic evolution as devised by Joseph Schumpeter fits in Witt's northwest quadrant of the matrix. I do not see any construct in Schumpeter's model that serves retention/heredity. I am leery of selection as a comprehensible mechanism in models of organizational evolution, particularly the Schumpeterian conception of the ecology surrounding the competition between New and Old Firms. There is too much aleatory uncertainty around the winners and losers to rely upon the outcomes as causally related to fitness, however defined. And co-evolutionary processes confound the selection mechanism at the population (non-community) level.The only part of the Universal Darwinist heuristics that I willingly embrace is variation. This is patently obvious, as innovation at the individual (firm) level is a purposeful exercise in creating variation from other individuals in the populations.

14.4Final Thoughts

In this paper, I used Ulrich Witt's (2008) categorizations of approaches to evolutionary economics as a foil to explore how generalizable evolutionary economics is; that is, can it reside comfortably with a Generalized Darwinian framework. I started with Witt's placement of Joseph Schumpeter's account of evolutionary economics with his abstraction that innovation within firms is the source of variation that initiates a process that leads away from equilibrium in some population context such as an industry or sector. Schumpeter then makes the idealization that the firm is characterized as a production function, a construct well understood in equilibrium economics. Innovation creates a New Firm, distinct in economic character from Old Firms -- the *sine qua non* of the static equilibrium economy. This idealization is more clearly stated in the 1939 account of *Business Cycles* than in the sources used by Witt. In this account, Schumpeter makes clear that the evolutionary process takes place in a complex ecological context he names the Economic Organism, a term embraced by David Sloan Wilson in a 2015 Evonomics blog article, "The Science is Clear. The Economy is an Organism", but not yet fully developed by him. In any case, outside of orthodox economics there exists a history of ecological accounts of organizations, competition, and complex inter-firm relationships that are tacit in *Business Cycles*.

As I update the Witt account of Schumpeter's evolutionary economics, complexity of the ecological context for innovation makes it difficult to match the blithe use of selection and retention as clear mechanisms for evolution of organizations and the markets in which they are embedded. Schumpeter ascribes this to derivative innovations in industries that are adjacent to the focal industry that appear as swarms in his account. Embedded mutualisms in this ecology can affect the competitive outcomes that are clearly abstracted in the category of Neo-Schumpeterian models. These latter accounts abstract from the ecosystem, focusing on a single industry with limited numbers of individual firms in the population. These accounts also rely on the construct of routines, a nearly insurmountable problem with their use as a mechanism within the models. I therefore find the possibility that Schumpeterian evolutionary economics can be considered as a case of Generalized Darwinism as defined by Witt (2008) and Hodgson and Knudsen (2010) to be unlikely.

Until Generalized Darwinism can accommodate the intricacies of economic co-evolution that occur in the complex ecology of adjacent markets, the evolutionary processes that are described by Schumpeterian innovation processes with limited diffusion, I cannot support the existence of Witt’s Schumpeter quadrant in his matrix. While certainly a non-monist account, this account or organization/market evolution as I have characterized it shares too few of the heuristics that Witt ascribes to his neighboring quadrants of Neo-Schumpeterian evolution and naturalistic diffusion.

I am more sanguine about the possibility that other evolutionary processes in the realm of economics can fit within the monistic category required of a Generalized Darwinism. As I pointed out, there is a class of entities that are not firms which may be a more comfortable fit. These include the inter-firm relationships that characterize markets. The types of transactions between firms (contracts, mutualisms like joint ventures and cooperatives) do not have the problems of inimitability that production functions -- including the accounts of the resource-based view (RBV) -- have. Those transactions are visible and are behavioral, rather than structural in a physical sense. Causal ambiguity and social complexity are modest, compared to RBV competitive strategy. These transactions may be much closer to the memes in cultural evolution, characterized more easily as diffused entities with identifiable selection criteria. Further investigation must wait for another paper.

To close, I propose that transaction entities can serve as populations that will not require significant auxiliary propositions to tie them to a Generalized Darwinian approach. But populations of firms that are more complex and larger than Willard Cochrane's cereal farms that are embedded in a complex ecology are not so easily equated with biological populations or cultural memes.

References

Alchian, A. A. (1953). Biological analogies in the theory of the firm: Comment. *American Economic Review*. 43(4), 600–603.

Barney, J.B. & Clark, D. (2007). *Resource-based theory: Creating and sustaining competitive advantage*. Oxford University Press.

Becker, M.C. (2004). Organizational routines: A review of the literature. *Industrial and Corporate Change*, 13(4), 643-677.

Becker, M.C., Lazaric, N., Nelson, R.R. & Winter, S.G. (2005). Applying organizational routines in understanding organizational change. *Industrial and Corporate Change*. 14(5), 775-791.

Carroll, G. R. (1988). *Ecological models of organizations.* Ballinger.

Cochrane, W.W. (1958) *Farm prices: Myth and reality*. University of Minnesota Press.

Delacroix, J., Swaminathan, A., & Solt, M. E. (1989). Density dependence versus population dynamics: An ecological study of failings in the California wine industry. *American Sociological Review*, 54, 245-262.

Eldredge, N. & Gould, S. J. (1972). Punctuated equilibria: An alternative to phyletic gradualism. In T.J.M. Schopf (Ed), *Models in paleobiology*. (pp. 82-115). Freeman, Cooper.

Hannan, M. T. & Freeman, J. (1977). The population ecology of organizations. *American Journal of Sociology,* 82(5), 929-964.

Hannan, M. T. & Freeman, J. (1989). *Organizational ecology*. Harvard University Press.

Hirshleifer, J. (1977). Economics from a biological viewpoint. *The Journal of Law & Economics*, 20(1), 1- 52.

Hodgson, G.M. & Knudsen, T. (2010). *Darwin's conjecture: The search for general principles of social and economic evolution*. University of Chicago Press.

Hutchinson, G.E. (1965). *The ecological theater and the evolutionary play*. Yale University Press.

Levallois, C. (2011). Why were biological analogies in economics 'a bad thing'? Edith Penrose's battles against social Darwinism and McCarthyism. *Science in Context*, 24, 465-485.

Nelson, R. R. & Winter, S. G. (1982a). The Schumpeterian tradeoff revisited*. American Economic Review,* 72(1), 114-132.

Nelson, R. R. & Winter, S. G. (1982b). *An evolutionary theory of economic change.* Harvard University Press.

Parmigiani, A. & Howard-Grenville. (2011). Routines revisited: Exploring the capabilities and practice perspectives. *The Academy of Management Annals*, 5(1), 413-453.

Penrose, E. T. (1952). Biological analogies in the theory of the firm. *American Economic Review*, 42(5), 804–819.

Penrose, E. T. (1953). Biological analogies in the theory of the firm: Rejoinder. *American Economic Review*, 43(4), 603–609.

Redlich, F. (1951). Innovation in business: A systematic presentation. *American Journal of Economics and Sociology*, 10(3), 285-291.

Reydon, T.A.C. & Scholz, M. (2009): 'Why organizational ecology is not a Darwinian research program', *Philosophy of the Social Sciences* 39, 408-439.

Ross, R. B. & Westgren, R. E. (2006). Economic returns to entrepreneurial behavior. *Journal of Agricultural and Applied Economics*, 30(2), 403-419.

Scholz, M. & Reydon, T.A.C. (2008). The population ecology programme in organization studies: Problems caused by unwarranted theory transfer, *Philosophy of Management* 6, 39-51.

Schumpeter, J. A. (1911). *Theorie der wirtschaftlichen Entwicklung*. Duncker & Humblot

Schumpeter J. A. (1934). *The theory of economic development*. Harvard University Press.

Schumpeter, J. A. (1939). *Business cycles: A theoretical, historical, and statistical analysis of the capitalist process.* McGraw-Hill.

Swaminathan, A. & Delacroix, J. (1991). Differentiation within an organizational form: Additional evidence from the wine industry. *Academy of Management Journal*, 34(3), 679-692.

USDA. 2020. *Farms and land in farms*. National Agricultural Statistics Service, Census of Agriculture.

Veblen, T. (1898). Why is economics not an evolutionary science? *Quarterly Journal of Economics,* 12, 373–397.

Vromen, J.J. (2006). Routines, genes, and program-based behavior. *Journal of Evolutionary Economics*, 16, 543-560.

Vromen. J.J. (2011). Routines as multi-level mechanisms. *Journal of Institutional Economics*, 7(2), 175-196.

Westgren, R. & Wuebker, R. (2019). An economic model of strategic entrepreneurship. *Strategic Entrepreneurship Journal*, 13(4), 507–528.

Witt, U. (2008). What is specific about evolutionary economics? *Journal of Evolutionary Economics*, 18, 547-575.

Chapter 15

Pluralism and epistemic goals: why the social sciences will (probably) not be synthesised by evolutionary theory

Simon Lohse[i]

Abstract

This article discusses Mesoudi et al.'s suggestion to synthesise the social sciences based on a theory of cultural evolution. In view of their proposal, I shall discuss two key questions. (I) Is their theory of cultural evolution a promising candidate to synthesise the social sciences? (II) What is the added value of evolutionary approaches for the social sciences? My aim is to highlight some hitherto underestimated challenges for transformative evolutionary approaches to the social sciences that come into view when one looks at these questions against the backdrop of actual scientific practice in the social sciences.

Keywords: cultural evolution, unity of science, quantification, idealisation, naturalism, interpretationalism.

---

[i] University of Lübeck ; E2LSI group at the Institute of Philosophy & Centre for Ethics and Law in the Life Sciences (CELLS), Leibniz University Hannover.

15.1Introduction

There is a long tradition of attempts to apply evolutionary thinking to the social sciences. In many cases, these attempts have been made by social scientists trying to use elements of evolutionary theory (broadly construed) for understanding the development of societies or institutional change. Herbert Spencer's theory of social evolution is arguably the most prominent classic (and notorious) example, but there are many more attempts to utilise evolutionary theory, for example in social systems theory (Luhmann, 2012, Chapter 3), in organisational theory (Hannan & Freeman, 1989), and in the context of economics and institutional theory (Hodgson & Knudsen, 2010).

In recent decades, there have also been attempts by scientists from other fields, in particular biology, psychology, and philosophy of science, to apply elements of evolutionary theory to the realm of the social sciences. These attempts are in most cases either led by the motivation to provide alternative – and: better or deeper – explanations of social phenomena, e.g. by providing an evolutionary mechanism that would explain the occurrence or persistence of a social institution (Alden Smith, 2000; Winterhalder & Smith, 2000) or to reduce human behavioural patterns to naturally evolved mechanisms of our brain (Cosmides & Tooby, 1989, 1994; Tooby & Cosmides, 1989). The other main motivation is to transform the social sciences based on evolutionary theory in order to improve their epistemic status in some sense, e.g. by aligning them to the more successful natural sciences. Alex Rosenberg is a prominent proponent of this approach. He argues for the reorganisation of the social sciences and their epistemic practices in light of the view that these "need to take seriously their status as a division of biology" (Rosenberg, 2017, p. 341). Although both motivations frequently overlap[1], there is a key difference. In the first case, the main idea is to supplement the social sciences with evolutionary theory or to provide evolutionary explanations for certain socio-cultural

---

[1] See, e.g., the ambiguous claims regarding sociology at the end of E. O. Wilson's chapter on "Man" in his *Sociobiology*,

phenomena. The second case aims at reconstructing the social sciences including their epistemic practices on a fundamental level according to or via a theory of cultural evolution.

In this article, I will focus on the latter motivation and discuss a prominent proposal to transform the social sciences that has been developed and defended by Alex Mesoudi together with Kevin Laland and Andrew Whiten. According to these authors, the social sciences can and indeed should be unified or "synthesised"[2] under the umbrella of a Darwinian theory of cultural evolution. In view of their proposal, I shall discuss two key questions. (I) Is their theory of cultural evolution a promising candidate to synthesise the social sciences? (II) What is the added value of evolutionary approaches for the social sciences? My aim is to highlight some hitherto underestimated challenges for transformative evolutionary approaches to the social sciences that come into view when one looks at these questions against the backdrop of actual scientific practice in sociology, cultural anthropology and political science, arguably centrepiece disciplines of the social sciences.

I will not reject an evolutionary synthesis of the social sciences based on fundamental ontological or epistemological objections against the application of evolutionary theory to the sphere of the social sciences. Hence, I will not be concerned with prominent themes in the literature, such as foundational criticism of mainstream concepts of culture in evolutionary theory (Lewens, 2012, 2015, p. 135f) and analyses of ontological dissimilarities between explananda in the biological and the social world (Reydon & Scholz, 2009; Reydon, 2021).Nor will I refer to discussions that criticise evolutionary explanations in the social sciences for their lack of explanatory power (Schatzki, 2001) or arbitrariness and reductionism (Dupré, 2001). Finally, I will not be focussing on problematic political implications of theories of cultural evolution (Chellappoo, 2021). Rather, I will use one of the most sophisticated candidates for transforming the social sciences based on a theory of cultural evolution as an illustrative

---

[2] I will use the terms 'synthesise', 'integrate' and 'unify' in the same sense.

example to draw out several pragmatic and methodological challenges for evolutionary approaches of this type. In doing so, I assume that many aspects of my discussion generalise to other "transformative projects".

This is how I will proceed. In section 2, I will introduce Mesoudi at al.'s core idea, namely to use evolutionary theory as a unifying theoretical framework for the social sciences. The main part of this article will scrutinise five central background assumptions of their approach (section 3). I will attempt to show that these assumptions cannot be taken for granted, and that each of them relies on a problematic, sometimes implicit rationale. The discussion will throw a sceptical light on the prospects of success for synthesising the social sciences through evolutionary theory (question I) and also address the question (II) after the added value of evolutionary approaches for the social sciences. I will conclude the article by highlighting implications of my discussion for the chances to integrate the social sciences and by making recommendations for increasing the likelihood of acceptance of evolutionary approaches in the social scientific community (section 4).

15.2 What does it mean to 'synthesise' the social sciences?

In this section, I will introduce the core idea of Mesoudi and colleagues and discuss their underlying motivation for synthesising the social sciences. I will base my discussion on the approach as it is developed in Mesoudi et al. (2006) and in Mesoudi (2011), building on the landmark work by Cavalli-Sforza/Feldman (1981) and Boyd/Richerson (1988; 2005). In a nutshell, their core idea is to use the theory of Darwinian evolution as a unifying theoretical framework for the multidisciplinary and multiparadigmatic social sciences, based on the observation that cultural evolution resembles biological evolution in key respects.[3] The proposed framework consists of (a) a common language for the social sciences based on

---

[3] I realise that "the theory of Darwinian evolution" is quite vague. As will become clear, however, the details of this notion do not matter for the discussion in this article.

evolutionary thinking, e.g. talk about populations, population-level patterns, and evolutionary histories as units of analysis; (b) a set of ideas concerning the concepts of culture and cultural change, e.g. ‘selective cultural transmission’ and ‘cultural drift’ as key mechanisms of cultural change; (c) postulates concerning (the right) explanatory tools and (the right) methodology, in particular setting quantification and mathematical evolutionary modelling as explanatory gold standard; and (d) a proposed structure of epistemic relations between fields of research (see fig. 1).

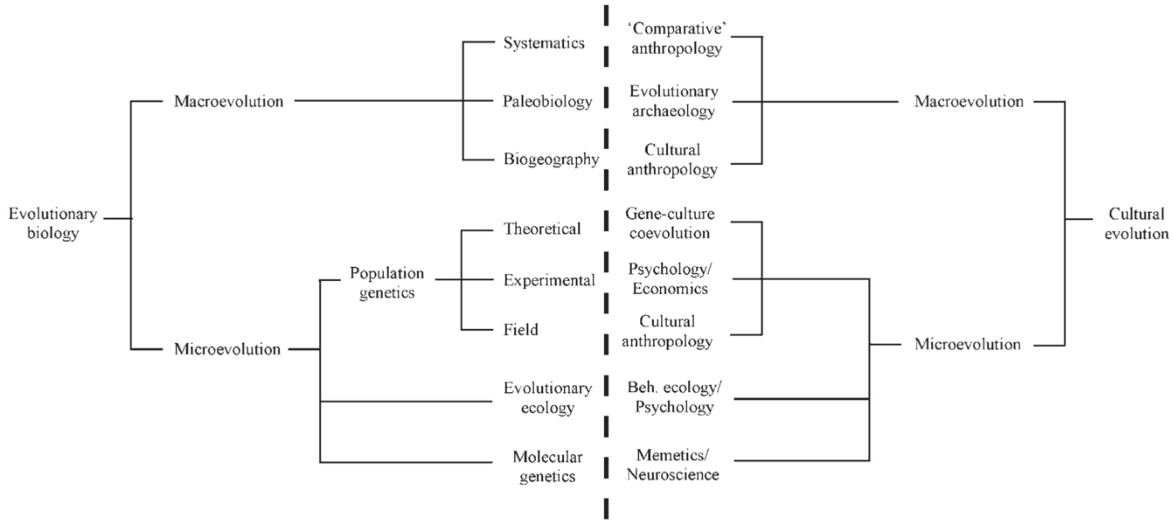


Figure 1: Epistemic relations between fields *(reprinted from Mesoudi, Whiten & Laland, 2006; there is a slightly modified version in Mesoudi, 2011, p. 211).*

The idea is to model the relationship between major disciplines and subdisciplines in the social sciences (broadly construed, i.e. including history/archaeology and, somewhat oddly, even parts of neuroscience) under the umbrella of a theory of cultural evolution and corresponding to epistemic relations between disciplines and subdisciplines in biology. For instance, as there are certain disciplines that aim at a macroevolutionary understanding of biological processes, there are supposedly corresponding social scientific disciplines that aim at a macroevolutionary understanding of cultural processes. The same goes for understanding microevolution. To illustrate the feasibility and fruitfulness of this idea, Mesoudi and colleagues discuss and compare several approaches in biology and the social sciences to point out analogies of their epistemic aims and practices in representing or explaining evolutionary processes. These examples include paleobiology characterised as “using the fossil record to identify prehistoric species and reconstruct their evolutionary history” (i.e. macrotrends)

(Mesoudi et al., 2006, p. 333), which they compare to archaeology's goal of analysing cultural artefacts to reconstruct their evolutionary history, such as the evolution of projectiles over a longer period of time. Another example compares mathematical models in biology that describe microevolutionary processes on the level of genes with models in socio-psychology that describe the transmission of cultural traits and believe frequencies in human populations (Mesoudi et al., 2006, p. 338).[4]

In these and other cases (see Mesoudi, 2011, Chapter 3-8), the main goal is to point out similarities between approaches in biology and the social sciences in order to make the case for the viability of an evolutionary synthesis of the social sciences in analogy to biology. The proposed epistemic framework is meant to serve as a kind of scaffold for this synthesis. The crux of this idea is to have explanatory projects situated at different levels (micro/macro) and with different foci (e.g. reconstructing lineages or distributions of cultural traits) that are organised around a common epistemic goal: understanding cultural evolution in a broadly Darwinian sense.

It is noteworthy that Mesoudi et al.'s approach does *not* aim at "biologising" the social sciences, i.e. it is not their intention to reduce everything social or cultural to the level of biology. Rather they explicitly argue for paying attention to important differences between biological and cultural evolution, as demonstrated by the consideration of the role that human agency and learning play in cultural evolution (Mesoudi, 2011, pp. 43-47). In other words, although their idea is to fundamentally transform the social sciences, their epistemic practices and relations – to synthesise them under the umbrella of a theory of cultural evolution – it is not about reducing them to or subsume them under the biological sciences.

---

[4] In Mesoudi (2011, pp. 212-216), there is an expanded discussion introducing disciplines that would have to be newly developed, such as ‚cultural evo-devo'.

The question remains, of course, what the rationale for this proposal is. Why do Mesoudi and colleagues think that the social sciences *should* be synthesised under the umbrella of a theory of cultural evolution? For one thing, their motivation seems to be in line with a tradition in Generalised Darwinism and (a substantially longer one) in philosophy of science that subscribes to the ideal of scientific unification based on general arguments for the advantages of theoretical parsimony and the assumption of metaphysical unity (see Aldrich et al., 2008 and Oppenheim & Putnam, 1958; for an overview of the debate in philosophy of science see Cat 2021). But there is also a more specific motivation (cf. Mesoudi et al., 2006, p. 330):

> Just as evolutionary theory served in the 1930s to synthesize the previously fractured biological sciences within a common and unifying theoretical framework, the interdisciplinary connections highlighted in the previous section [where Mesoudi discusses examples of evolutionary explanations in different social science disciplines, S.L.] suggest that a similar synthesis is possible and may be nearing for the social sciences. (Mesoudi, 2011, p. 210)

This point is derived from a diagnosis of the social sciences as being disconnected from each other and fragmented within – and: a conceivable solution for this issue that is modelled after the "evolutionary synthesis" in biology. As the evolutionary synthesis has provided cohesion and integration for biology[5], it can, according to Mesoudi et al., do the same for the social sciences. It is supposed to bridge different approaches within disciplines and between disciplines in the social sciences that focus either on the microlevel *or* the macrolevel of culture and cultural change (Mesoudi, 2011, pp. 51–53). Through this kind of synthesis, Mesoudi and colleagues believe, there will be much more communication, cohesion and progress in the social sciences than is possible now, as there will be more epistemic integration and a more complex understanding of the different aspects and effects of culture.

---

[5] This assessment of the role of the evolutionary synthesis in biology is not uncontroversial (Lewens, 2012, p. 463). However, I will take it for granted in this paper.

Prima facie, this might be considered a plausible rationale for an evolutionary synthesis of the social sciences. However, there remains a curious fact. Consider the current state of sociology, cultural anthropology and political science. All three can indeed by characterised as multiparadigmatic. But this is not so because there have been no attempts to integrate these disciplines. On the contrary, there have been an abundance of "paradigm synthesis approaches" in and between these disciplines, including Parsons (Parsons, 1937; Parsons & Shils, 1951), Giddens (1984), Coleman (1990), Esser (1993), and Mayntz/Scharpf (1995). Not one of them has been successful, at least not in a sustainable way (I will come back to this point below). Why is that? While I do not have a complete theory that can answer this question, I can provide a starting point for an adequate answer. There has not been any successful synthesis of the social sciences, as the reasons for the pluralistic structure of these disciplines have not been thoroughly analysed. Synthesising approaches typically make a number of assumptions concerning said reasons and start from there, despite the fact that these assumptions may not necessarily be very well backed up by empirical evidence (Lohse, 2017a). As I will be showing in the next section, using sociology, cultural anthropology and political science as examples, this holds true for Mesoudi and colleagues too. Their proposal, including its main rationale, to synthesise the social sciences in analogy with the biological sciences is based on several background assumptions concerning the social sciences that cannot be taken for granted but are problematic to varying degrees.

15.3Background assumptions and challenges

There are five background assumption of the proposal to synthesise the social sciences that I will discuss in this section (based on Mesoudi et al., 2006 and Mesoudi, 2011, Chapter 1 & 10). I will attempt to show that these assumptions rest, to different degrees, on unstable ground.

Assumption (1) The social sciences are fragmented, which is a main reason for their lack of progress.

This assumption is an important aspect of the rationale to synthesise the social sciences and is expressed in Mesoudi et al.'s dissatisfaction with the epistemic status of the social sciences. The social sciences are characterised as lacking empirical success as compared to the natural sciences, in particular in terms of precise models and widely accepted explanations of cultural phenomena and cultural change. This is attributed to a large extent to the epistemic fragmentation of the social sciences in different disciplines, subdisciplines and paradigms (Mesoudi, 2011, p. 208).

Assumption (1) faces several difficulties. For one thing, it fails to take adequately into account alternative explanations for the lack of social scientific progress. Let us grant that there is more cohesion and epistemic success in the natural sciences than in the social sciences. They provide us with relatively stable explanatory frameworks as well as more exact descriptions and predictions of phenomena than the social sciences. This does not necessarily mean that it is their cohesion that is the main reason for their success or that it is the fragmentation of the social sciences that is responsible for insufficient progress. Indeed, there are many hypotheses in the literature that attempt to explain the difference in epistemic success. Some think that it is the sheer complexity of social systems that makes it extremely hard for the social sciences to describe or predict them in adequate detail (Scriven, 1994). Maybe the lack of progress can be explained with reference to human agency, which may not be expressible in causal laws (Tanney, 2013). Rosenberg (2012, pp. 14ff) argues that methodological and practical challenges are among the main reasons for the relative lack of epistemic success in the social sciences. For instance, experiments with human subjects are in many cases not feasible for ethical reasons or because they would distort their results through observer-expectancy effects. *Prima facie*, all of these explanations seem plausible. However, without much more detailed analyses of the epistemic status quo in the social sciences, it seems hard to decide which of these are actually on the right track.

There is an additional issue. There are epistemological reasons that make it seem questionable to just *assume* that epistemic fragmentation is a problem for the social sciences. Maybe we should understand the social sciences as a genuinely pluralistic enterprise that investigates and highlights different aspects of a complex cultural and social reality. This could be an epistemically fruitful approach, as it might avoid a myopic scientific understanding of said reality.[6] Pluralism could also be advisable from a methodological point of view. Triangulating different perspectives and approaches could for example alleviate the shortcomings of each perspective. (for a more detailed account along these lines see Feyerabend, 1978, 1999; Lloyd, 1997; also see Kellert et al., 2006). This could be particularly important in case of the social sciences, as implicit and hard to detect biases and values might be influencing the scientific investigation of human affairs to a greater degree than in the natural sciences.

At this point, Mesoudi and colleagues could object. They do not characterise the social sciences as pluralistic, but as *fragmented*. There is a difference. They point out that there is a lack of integration and communication between different disciplines and between micro and macro approaches within disciplines, which is precisely the issue here. According to Mesoudi this shows in at least two ways. First, there is insufficient accumulation of knowledge in the social sciences. The reason for this is – and I believe this is spot on (Mesoudi, 2011, p. xiii) – a tendency to re-invent the wheel in different, disconnected disciplines (think of the "discovery" of the influence of values and norms in empirically informed economics). Second, Mesoudi diagnoses that there is a lack of methodological and conceptual exchange between disciplines, again being problematic for progress in the social sciences:

> The traditional social sciences are hindered by the separation of different methods and different subjects into different disciplines: while psychologists conduct laboratory experiments, cultural anthropologists conduct ethnographic fieldwork, archaeologists document prehistory, and economists construct models of market systems. (Mesoudi, 2011, p. 208)

---

[6] Accordingly, Kneer and Moebius (2010) argue that the multiparadigmatic state of sociology should be seen as a sign of its prosperity and controversy as a key element of its disciplinary constitution.

However, Mesoudi overstates his case here. For one thing, there is conceptual exchange going on in the social sciences as the many papers and books with overviews and comparisons of key concepts and theories suggest (e.g. Groh, 2019; López & Scott, 2000). This is also clear from the existence of overlap in fields of study, say in educational science and sociology of education, and from the existence of hybrid fields such as political sociology. In addition, it is misleading to characterise different social sciences as characterised through a dominant methodology. Experiments and ethnographic methods thrive in sociology, as do economic models in social choice theory and interview studies in economics. Thus, there is definitely conceptual and methodological exchange between the different social sciences. The social sciences are not as fragmented as Mesoudi sees them and that makes assumption (1) seem doubtful. It would be an overstatement, however, to say that there is *no* fragmentation in the social sciences. Mesoudi and colleagues do have a point. There certainty could be more exchange and knowledge integration in the social sciences, less re-inventing the epistemic wheel and less unnecessary fights between disciplines (e.g. economics vs. sociology) and paradigms (e.g. rational choice theory vs. practice theory) and this might indeed be fruitful for progress in the social sciences.

Assumption (2) The fragmented state of the social sciences is mainly due to the unavailability of an integrative theoretical framework, such as evolutionary theory.

With the mentioned caveats in mind, it seems fair so say that the social sciences are indeed fragmented *to some degree*. As already suggested in section 2, one of the main reasons for this state of affairs is, according to Mesoudi et al., the unavailability of an integrative framework, in particular a framework that would be able to bridge different disciplines and micro/macro approaches in the social sciences. The latter aspect is important as the social sciences are seen by Mesoudi and colleagues as particularly deeply divided in micro/macro approaches.

However, this view of the social sciences seems to exaggerate the issue. Although Mesoudi et al.'s assessment may hold true for certain subfields in the social sciences, the micro/macro divide is not anymore the separating line that is used to be 40 years ago. There are now many frameworks in the social sciences that actively promote micro/macro integration. This includes rational choice theory (Coleman, 1990), practice theory (Bourdieu, 1977; Bourdieu & Wacquant, 1992) and certain authors in Analytical Sociology (see Demeulenaere, 2011). But there is an even bigger problem. Let us assume for a moment that there was indeed a scarcity of integrative micro/macro frameworks in the social sciences. Why should we assume that this fact would be the main reason for the fragmented state of the social sciences? There seems to be a lack of evidence to support this assumption.

As a matter of fact, we do not seem to know what the underlying reasons for the multiparadigmatic/fragmented state of social sciences actually is. To be sure, there exist several hypothetical explanations that have been suggested in the literature. Some think that the subject area of the social sciences is too flexible and changes too fast to develop stable explanatory frameworks that capture more than a fragment of cultural reality (see the discussion in McIntyre, 1993). Others have highlighted the complexity of the social world as a key factor (see above). Maybe the social sciences can never hope to have more than partial explanatory frameworks for some aspects of socio-cultural reality. Thomas Kuhn (2000[1991]) has suggested yet another possibility. He draws attention to the hermeneutical nature of the social sciences as a reason for their multiparadigmatic state. According to Kuhn, the social sciences constantly redescribe and reinterpret social reality which makes it hard to enter a state of normal science. There are many more hypotheses of why the social sciences are fragmented (to some extent), including sociological hypotheses hinting at the strong tendency for building schools as distinguishing brands (Schimank, 2012). But this is all these are: hypothesis of *possible* explanations for the status quo. At present, we do not have a corroborated explanation of the multiparadigmatic state of the social sciences but only a number of competing explanatory hypothesis.[7] An important consequence of this is that it is unclear what obstacles an integrative

---

[7] A main reason for this epistemic gap is that there is, to my knowledge, no research programme in history and philosophy of science that investigates this question.

approach has to contend with and to what extent a unifying framework in terms of evolutionary theory (or otherwise) is even possible at present.

Assumption (3) All of the social sciences investigate the same "cultural stuff".

This assumption is the ontological core of the synthesising project. Since all of the social sciences investigate different aspects of culture at different levels and in different ways, they can be integrated through a theory of cultural evolution. So what is 'culture'? Mesoudi and colleagues provide us with a broad characterisation of culture that is based on earlier work in cultural evolution theory and meant to be all-encompassing:

> Following Richerson and Boyd (2005), we define culture as 'information capable of affecting individuals' behaviour that they acquire from other members of their species through teaching, imitation, and other forms of social transmission' (p. 5). 'Information' is employed as a broad term incorporating ideas, knowledge, beliefs, values, skills, and attitudes. (Mesoudi et al., 2006, p. 331)

While the paradigm case of culture seems to be "information in the head", the concept is broader. In his book on cultural evolution, Mesoudi elaborates:

> Whereas genetic information is stored in sequences of DNA base pairs, culturally transmitted information is stored in the brain […] as well as in extrasomatic codes such as written language, binary computer code, and musical notation. And whereas genetic information is expressed as proteins and ultimately physical structures such as limbs and eyes, culturally acquired information is expressed in the form of behavior, speech, artefacts, and institutions. (Mesoudi, 2011, 3)

This conceptualisation of culture may raise some eyebrows. How apt is the analogy of culture and genetic information? Can skills such as riding a bike really be considered as information encoded (solely?) in neural patterns? Do patterns of behaviour, artefacts and institutions *encode* or *express* cultural information? Both?

These conceptual questions point to legitimate concerns. However, I want to focus on another issue, namely whether social scientists are really talking about the same thing, when they talk about 'culture'. This does not seem to be the case. Rather, there are a vast number of ideas in the social sciences concerning the right way to characterise what culture actually means (Sewell, 2005, Chapter 5). Smith (2016) lists many different definitions that have been offered in the social sciences, characterising culture as *inter alia* ideas, values, beliefs, meaning, symbolic codes, mental representations, discourses, semiotic systems, artefacts, actions, social processes, practices and various combinations thereof. Smith highlights that 'culture' is an extremely contested concept which is characterised as vague and even incoherent. Note that this is not just a quibble. Different characterisations of culture are not merely highlighting different aspects that could easily be reconciled using the definition of culture as mentally realised information and expressions of this information in behaviour and artefacts. Rather, they point to different and partially incommensurable social ontologies of culture that exist in different social scientific schools. While some social scientists see culture as a mental phenomenon, others think of it as a structure "out there", as an implicit systems of rules for behaviour or a network of meaning (see the landmark discussion in Geertz, 1973). These ideas cannot easily be reconciled with each other. They have been the subject of long lasting debates in the social sciences that can also be observed regarding other basic concepts, e.g. 'institutions' and 'organisations'. Moreover, these different conceptions of culture are interwoven with the core ontological assumptions of different paradigms about the nature of socio-cultural reality and specific explanatory preferences (e.g. a preference for reductionist explanations in individualism). An implication of this is that these conceptions of culture cannot be integrated without considering their theoretical embeddedness.

A synthesising approach like Mesoudi et al.'s would have to engage productively with the fact that there are currently many different explanatory frameworks with many different conceptions concerning the basic entities and processes in socio-cultural reality. The alternative would mean either to be only able to integrate those approaches that are already aligned with a more or less particularistic picture of culture and microfoundational approaches, such as methodological individualism (see Lewens, 2012, 2015, p.139) – this is the best case scenario, or to merely establish another paradigm with its own core ideas regarding what culture is *next* to the already existing, well-established ones.

Assumption (4) Social scientists (except economists and psychologists) reject simplifying quantification and mathematical modelling without good reasons.

Although this is not a motivating assumption for Mesoudi and colleagues, it is important for making their case. Assumption (4) points to what they see as a major but ultimately irrational obstacle for evolutionary approaches to cultural reality; an obstacle that needs to be overcome by the social sciences to become "truly" scientific (Mesoudi et al., 2006, p. 337). As stated above, Mesoudi et al. believe that quantification and mathematical evolutionary modelling – which presuppose idealisation, in particular in form of simplifications that allow formal descriptions of complex phenomena – are key elements that an evolutionary approach can bring to the table. It is this methodological gold standard that they want to make accessible for the social sciences through an evolutionary synthesis. The main reason Mesoudi provides for this aim is that quantification and mathematical modelling enable more precise descriptions of cultural phenomena and processes. In addition, they allow for formal testing of hypotheses against real world data, thereby surpassing the possibilities of "verbal arguments back and forth between scholars, each of whom believes their pet theory to be better, with no real way to determine who is correct" (Mesoudi, 2011, p. 206). In this sense, quantification and simplifying mathematical modelling of cultural phenomena and processes are regarded as more rigorous and *ipso facto* more scientific than non-quantitative approaches (cf. Mesoudi et al., 2006, pp. 329f; Mesoudi, 2011, p. 205).

It is true that social scientists sometimes reject quantification and mathematical modelling for the wrong reasons. They may criticise formal approaches in very general terms as a simplistic reduction of culture that does not do justice to its complexity, without acknowledging the advantages that idealisation, i.e. simplifying reduction of complexity, can have for promoting our understanding of the world (see Potochnik, 2017).[8] There are, however, two problems with assumption (4).

The first problem is that it is an extreme overgeneralisation. Many parts of the social sciences, including sociology, cultural anthropology and political science, are decidedly quantitative and use idealising mathematical modelling. This includes agent-based modelling, often in cooperation with scientists from other fields, to explore such different things as voting behaviour (Fowler & Smirnov, 2005) and racial disparities in incarceration rates (Lum et al., 2014) as well as the use of mathematical game theory models to better understand territorial conflicts in international relations (Carter, 2010). In fact, much research in leading journals in sociology and political science relies on quantitative approaches using sophisticated statistical regression methods, computer simulations and other quantitative tools (an abundance of examples can be found in *The American Journal of Political Science, The European Journal of Sociology* and in many other top journals in the respective fields). Cultural anthropologists, while often relying on ethnographic observations, also regularly use – and have for a long time – quantified methods (see the widely used textbook *Research Methods in Anthropology: Qualitative and Quantitative Approaches*, Bernard, 2017). Furthermore, there exist associations for computational social sciences as well as journals such as the *Journal of Artificial Societies and Social Simulation* and *The Journal for Mathematical Sociology* (since 1971!). In light of this, it is misleading to say that most social sciences reject simplification, quantification and mathematical modelling.

The second problem for assumption (4) is that quantitative approaches in the social sciences are in many cases rejected with *good* reasons. Many social scientists and philosophers of science (including myself) believe that the social sciences are - and indeed need to be - an

---

[8] The same goes for the sometimes sweeping criticisms of evolutionary approaches to socio-cultural phenomena as necessarily biologistic and wrong-headed.

interpretive enterprise to a significant part. This does not mean that quantification and mathematical models are useless or that qualitative research projects cannot benefit from quantitative approaches (as the methodological trend of mixed methods research designs shows, see Creswell & Plano Clark, 2011). Rather, the idea of the social sciences as an interpretative enterprise means that there is a significant part of it that has to rely on hermeneutics (see the above remarks on Kuhn), i.e. understanding and sense-making of socio-cultural reality. This is no deficit of the social sciences and it does not make these parts of the social science any less scientific. It merely reflects the meaning dimension of socio-cultural reality and the centre stage that the concept of intentional agency understood as an interpretable phenomenon takes in the social sciences. In fact, much research in the social sciences is about understanding the meaning of social practices (e.g. in ethnographic research), discovering layers of sometimes latent meanings behind certain acts (e.g. qualitative research in the Mannheimian tradition), and the critical reconstruction of ideological assumptions underlying cultural institutions (e.g. in feminist political theory). There seems to be no good reason for assuming that research programmes along these lines can be *replaced* with formal models or need to be overcome completely through quantitative approaches. On the contrary, there is a long tradition emphasising that the social sciences, understood as an partially interpretative enterprise, can in an important sense provide a deeper explanation of their subject matter than the natural sciences. Max Weber offers what remains one of the best justifications for this claim. He demonstrates that an explanation in the social sciences can go beyond the determination of causal mechanisms or the mere description of regularities. For this purpose, intentional actions and social practices that underlie socio-cultural phenomena must be placed in a context of meaning – i.e. in a context of other meaningful actions, practices and intentions – and thus made truly intelligible for us (Weber, 1978[1921/22], Chapter 1). This is analogous to texts that we interpret and understand with reference to other texts, and often also the assumed intentions of the author (Taylor, 1971).

Any approach that aims at synthesising the social sciences will need to integrate this hermeneutic tradition and show how it can be fruitfully combined with quantitative approaches. Arguing for the inferiority of qualitative research, as Mesoudi and colleagues do risks downplaying the richness of human agency and contributing to the long lasting and paralysing qualitative/quantitative controversies in the social sciences. To put in another way, it is not true that a discipline can only be “fully scientific” if it is quantitative and makes testable predictions (Mesoudi, 2011, pp. 18ff, 205) – rather, this assumption implies an untenable scientific monism. If recent developments in philosophy of science have shown anything, it is this. There is not *the* scientific approach, but many different approaches in the sciences, and this includes hermeneutics as a systematic and more rigorous form of everyday interpretation of the socio-cultural world around us (Hoyningen-Huene, 2013, pp. 71–77).

As a side note, it does not help Mesoudi et al.’s case to label non quantitative approaches anti-naturalistic or lax and throw them out with the post-modern bathwater (see Mesoudi 2011, p. 19f for a misleading characterisation of the hermeneutical approach in anthropology; see also Mesoudi et al., 2010). Hermeneutical approaches in the social sciences are in most cases not anti-science or postmodernist, but adhere to their own methodological standards and rules – traceability, reflexivity, coherence of interpretation etc. – as even a brief look at the methodological literature in qualitative social research clearly shows. Quantification is not the only way to being rigorous.

Assumption (5) The social sciences share an epistemic core goal: explaining cultural change and the effects of culture on human behaviour.

This is an essential background conviction of Mesoudi and colleagues. Without this assumption, it would not make sense to propose a theory of cultural evolution as epistemic core of the social sciences. A theory of cultural evolution is assumed to be able to integrate the social sciences precisely because it can organise them in alignment with a shared epistemic goal, namely explaining cultural change and the effects of culture on human behaviour.

The problem with assumption (5) is that it neglects the great diversity of epistemic goals of the social sciences and the question of what added value evolutionary theory and modelling have for achieving these goals.[9] According to Mesoudi and colleagues, evolutionary approaches to culture are much better than traditional social scientific methods, as they allow for more precise modelling of socio-cultural phenomena and processes. They substantiate their argument with many impressive examples for this claim. The question remains, of course, whether this means that evolutionary approaches are preferable *tout court.* I have already expressed some concerns about this above, but more can be said when considering the assumed added value of evolutionary approaches in light of different epistemic goals in the social sciences.

The added value of evolutionary approaches does not exist in a vacuum. It rather needs to be determined in light of the epistemic goals pursued in each given case. There are, of course, several epistemic projects pursued by social scientists that are well aligned with Mesoudi's and colleagues' interest to describe transmission, change and diffusion of cultural phenomena and explain the difference that culture makes for human behaviour. Many social scientists are interested in related issues. Accordingly, evolutionary models of culture can be helpful for understanding general social mechanisms for observational learning (e.g. based on prestige bias) and they can enrich individualist explanations of patterns of political change (e.g. through transmission chain experiments). However, these and similar cases are by far not the only epistemic games in town. There are many more projects in the social sciences. These range from the description of side-effects of policy and the latent properties of political systems to ethnographic explorations of social spaces and the deconstruction of social categories such as race or gender (see also the above discussion of hermeneutical projects). Consider these examples, chosen from leading social science journals:

- In "Intended and Unintended Effects of the War on Poverty: What Research Tells Us and Implications for Policy", the authors review and evaluate the evidence on causal effects and side effects of policy programmes in the US to reduce poverty (Bitler & Karoly, 2015). One of

---

[9] This part of the article draws on work on explanatory pluralism by van Bouwel & Weber (2008a, 2008b).

their primary epistemic goals is to discover and better understand – often unintended and latent – side effects of policy measures, e.g. on employment rates, in order to inform and improve policy-making.

- In two highly cited papers in *American Political Science Review* and *Journal of Conflict Resolution*, Maoz and his co-authors deploy a comparative research design to explore the question what regime attributes might be related to the likelihood of entering into a war with another country (e.g. are democracies less likely to enter a war with each other?) (Maoz & Abdolali, 1989; Maoz & Russett, 1993).

- In a paper in *American Sociological Review*, Smith (2014) reports the results of a long-term ethnographic study that aims to explore how ethnic identity changes through the life course of individuals and in what ways these changes need to be situated in historical and institutional context.

- In "Gender and the Career Choice Process: The Role of Biased Self-Assessments", Correl (2001) reports the results of a quantitative study investigating the exact role that negative self-ascription of mathematical abilities by women plays in their career paths, at the same time helping to further debunk the view that biological differences are determining mathematical abilities in men and women.

The issue for Mesoudi et al.'s proposal that come into view through these examples is this: These are no cherry-picked cases or exceptional research projects.[10] Rather they are sophisticated and successful examples of *typical* research projects in the social sciences that are no more and no less in the epistemic centre of the social sciences than explaining culture and cultural evolution. (I do not think that we have any grounds to think that there *is* a central epistemic goal of the social sciences.) At the same time, it is not clear to what extent these

---

[10] In fact, there might be a sampling bias in Mesoudi et al.'s examples. In discussing the benefits of a theory of cultural evolution for the social sciences, they tend to discuss examples from anthropology, psychology, behavioural economics and other subfields that are already aligned with their focus on cultural change, transmission and explaining general patterns of human behaviour.

projects and many other projects in the social sciences would benefit from a theory of cultural evolution or evolutionary modelling. Mesoudi and colleagues provide no reasons to think otherwise. In fact, it would be necessary to show to what extent evolutionary thinking could be useful for projects with these or similar epistemic goals, e.g. by showing that certain claims or explanations are wrong. This is not denying that an evolutionary approach can be an extremely useful tool – but it is but one tool next to others and no better tool for any purpose (cf. Lewens, 2015, p. 146). The adequacy of a theory, model or method depends on the given epistemic goal to be pursued, which may or may not be aligned with what evolutionary approaches can offer.[11] And even in cases where epistemic goal and evolutionary tool are well aligned, evolutionary approaches may not provide the best *available* explanations, at least not without embedding these within the rich explanatory resources of sociology, cultural anthropology etc. This is precisely what has been at the centre of much criticism of evolutionary explanations of socio-cultural phenomena that highlight observational learning and transmission chains, while neglecting the role of the institutional environment as key explanatory factor. Against this backdrop, it seems unclear why cultural evolution theory and evolutionary modelling should be at the epistemic centre of a synthetic framework for the social sciences.

15.4 Conclusion

Let us take stock. I have shown that assumption (1) does not adequately account for alternative explanations for the (relative) lack of social scientific progress and the (potential) benefits of having a plurality of perspectives in the social sciences. This seems to mitigate the strength of the assumption. A major problem for assumption (2) is a lack of evidence to back it up, pointing to a more general problem: At present, we do not seem to have a corroborated theory that can explain the multiparadigmatic state of the social sciences. Assumption (3) underestimates the extent to which different and partially incommensurable social ontologies

---

[11] Sober (1992) makes a related point by arguing that social scientists are more interested in the sources of transmission systems than in their empirical consequences, which is why evolutionary models may not be very useful in many cases. This also points to different epistemic goals in the social sciences.

of culture exist in the social sciences – this is a tough challenge for every synthetic approach to the social sciences. Assumption (4) addresses a potential (if exaggerated) barrier for Mesoudi et al.'s synthetic approach but neglects the hermeneutical dimension of many research projects in the social sciences as a good reason for resisting (too much) quantification. Assumption (5) underestimates the wide variety of epistemic goals of social scientists. This leads to (a) overstating the case for the potential of a theory of cultural evolution to serve as the epistemic core of the social sciences and (b) overstating the usefulness of evolutionary models for the social sciences.

What is the upshot of this for Mesoudi et al.'s synthesising approach to the social sciences? While some of the identified issues might be mitigated by modifying and expanding the proposed framework, others, namely issues connected to ontological incommensurability, hermeneutics and goal pluralism, will prove to be extremely thorny. These touch on the foundations of the social sciences and will be difficult (if not impossible) to overcome with the proposed approach.

The concurrence of the discussed problems seriously undermines the prospects of success for Mesoudi et al.'s project – and I believe for similar projects as well. It is the extensive and deep-cutting pluralism in the social sciences, including their ontology, methodology and epistemic goals, that presents synthesising projects with major obstacles. Even if we think that there *should* be more integration and less pluralism in the social sciences (and this is by no means uncontroversial, see above), it seems unlikely that we actually *can* synthesise the social sciences, at least not in the foreseeable future.

After this rather pessimistic assessment, let me conclude with two constructive notes. The first note aims at scientists and philosophers that are interested in promoting "the integrative project". To make progress they would need to invest more time in understanding the challenges for a synthetic approach to the social sciences. This means paying much more attention to actual research practices in the social sciences in order to gain a proper understanding of their pluralistic state and to find out what it would take for a synthetic project to succeed: *Bottom-up instead of top-down transformation*. This may, of course (pessimist again), lead to the conclusion that integration will remain impossible, e.g. due to irresolvable ontological or methodological incompatibilities. But it could also lead to ways to alleviate some of the identified problems, e.g. by showing that there is more ontological compatibility in the social sciences than one would think (Lohse, 2017b, 2019).

The second note addresses the acceptance of evolutionary approaches in the social sciences. It is important for proponents of evolutionary approaches to make the case for their usefulness in a way that is recognised by social scientists. Advocates of evolutionary approaches need to show why their approach is better in light of the epistemic goals of sociologists, political scientists etc. In other words, it is not helpful to shift the burden of proof like this if one wants to gain acceptance in mainstream social science:

> We maintain that critics [of evolutionary approaches] must empirically demonstrate that the existence of intent does in fact invalidate an evolutionary account of human culture […]. (Mesoudi et al., 2006, p. 345)

It is the other way around. It has to be demonstrated how evolutionary theory and modelling can help answering questions that are of interest in the social sciences. *Nota bene*, Mesoudi and colleagues frequently, though not always, pursue this avenue. But in many cases, especially in the context of Generalised Darwinism, there is a top-down perspective that places unification and generalisability above all other epistemic goals, values (such as explanatory depth) and aspects of social scientific practise (see Chellappoo, 2021 for a more detailed analysis of this point). Approaches of this kind will continue to be seen – and rightly so – as an attempt at scientific imperialism.

References

Alden Smith, E. (2000). Three Styles in the Evolutionary Analysis of Human Behavior. In L. Cronk, N. Chagnon, & W. Irons (Eds.), *Adaptation and Human Behavior: An Anthropological Perspective* (pp. 27–46). Aldine de Gruyter.

Aldrich, H. E., Hodgson, G. M., Hull, D. L., Knudsen, T., Mokyr, J., & Vanberg, V. J. (2008). In defence of generalized Darwinism. *Journal of Evolutionary Economics*, *18*(5), 577–596. https://doi.org/10.1007/s00191-008-0110-z

Bernard, H. R. (2017). *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. Rowman & Littlefield.

Bitler, M. P., & Karoly, L. A. (2015). Intended and Unintended Effects of the War on Poverty: What Research Tells Us and Implications for Policy. *Journal of Policy Analysis and Management*, *34*(3), 639–696. https://doi.org/10.1002/pam.21842

Bourdieu, P. (1977). *Outline of a Theory of Practice*. Cambridge University Press.

Bourdieu, P., & Wacquant, L. J. D. (1992). *An Invitation to Reflexive Sociology*. University of Chicago Press.

Boyd, R., & Richerson, P. J. (1988). *Culture and the Evolutionary Process*. University of Chicago Press.

Carter, D. B. (2010). The Strategy of Territorial Conflict. *American Journal of Political Science*, *54*(4), 969–987. https://doi.org/10.1111/j.1540-5907.2010.00471.x

Cat, J. (2021). The Unity of Science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2021 edition)*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/scientific-unity/

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton University Press.

Chellappoo, A. (2021). Cultural evolution: A case study in global epistemologies of science. In D. Ludwig, I. Koskinen, Z. Mncube, P. Luana, & L. Reyes-Galindo (Eds.), *Global Epistemologies and Philosophies of Science* (pp. 208–219). Routledge.

Coleman, J. S. (1990). *Foundations of Social Theory*. Belknap Press.

Correll, S. J. (2001). Gender and the Career Choice Process: The Role of Biased Self-Assessments. *American Journal of Sociology*, *106*(6), 1691–1730. https://doi.org/10.1086/321299

Cosmides, L., & Tooby, J. (1989). Evolutionary Psychology and the Generation of Culture, Part II: Case Study: A Computational Theory of Social Exchange. *Ethology and Sociobiology*, *10*(1–3), 51–97. https://doi.org/10.1016/0162-3095(89)90013-7

Cosmides, L., & Tooby, J. (1994). Better Than Rational: Evolutionary Psychology and the Invisible Hand. *American Economic Review*, *84*(2), 327–332.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research* (2nd ed.). SAGE.

Demeulenaere, P. (Ed.). (2011). *Analytical Sociology and Social Mechanisms*. Cambridge University Press.

Dupré, J. (2001). *Human Nature and the Limits of Science*. Clarendon Press.

Esser, H. (1993). *Soziologie. Allgemeine Grundlagen*. Campus.

Feyerabend, P. (1978). *Science in a Free Society*. NLB.

Feyerabend, P. (1999). Outline of a Pluralistic Theory of Knowledge and Action. In J. Preston (Ed.), *Knowledge, Science, and Relativism. Philosophical Papers Vol. 3* (pp. 104–111). Cambridge University Press.

Fowler, J. H., & Smirnov, O. (2005). Dynamic Parties and Social Turnout: An Agent-Based Model. *American Journal of Sociology*, *110*(4), 1070–1094. https://doi.org/10.1086/426554

Geertz, C. (1973). *The interpretation of cultures: Selected essays*. Basic Books.

Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Polity Press.

Groh, A. (2019). *Theories of Culture*. Routledge.

Hannan, M. T., & Freeman, J. (1989). *Organizational Ecology*. Harvard University Press.

Hodgson, G. M., & Knudsen, T. (2010). *Darwin's conjecture: The search for general principles of social and economic evolution*. University of Chicago Press.

Hoyningen-Huene, P. (2013). *Systematicity: The nature of science*. Oxford University Press.

Kellert, S. H., Longino, H. E., & Waters, C. K. (Eds.). (2006). *Scientific Pluralism (Minnesota Studies in the Philosophy of Science)*. University of Minnesota Press.

Kneer, G., & Moebius, S. (Eds.). (2010). *Soziologische Kontroversen: Beiträge zu einer anderen Geschichte der Wissenschaft vom Sozialen* (1st ed.). Suhrkamp.

Kuhn, T. S. (1991). The Natural and the Human Sciences. In T. S. Kuhn & J. Conant (Eds.), *The Road Since Structure Philosophical Essays, 1970-1997 with an Autobiographical Interview* (pp. 216–223). University of Chicago Press (2000).

Lewens, T. (2012). Cultural evolution: Integration and skepticism. In H. Kincaid (Ed.), *The Oxford Handbook of Philosophy of Social Science* (pp. 458–480). Oxford: Oxford University Press.

Lewens, T. (2015). *Cultural Evolution: Conceptual Challenges*. Oxford University Press.

Lloyd, E. A. (1997). Feyerabend, Mill, and Pluralism. *Philosophy of Science*, *64*, S396–S407.

Lohse, S. (2017a). Die Multiparadigmatik der Soziologie als Erklärungsgegenstand einer integrierten Wissenschaftsforschung. *Theoretische Soziologie*, *6*(2), 237–246.

Lohse, S. (2017b). Pragmatism, Ontology, and Philosophy of the Social Sciences in Practice. *Philosophy of the Social Sciences*, *47*(1), 3–27. https://doi.org/10.1177/0048393116654869

Lohse, S. (2019). *Die Eigenständigkeit des Sozialen. Zur ontologischen Kritik des Individualismus*. Mohr Siebeck.

López, J., & Scott, J. (2000). *Social Structure: Concepts in the Social Sciences*. Open University Press.

Luhmann, N. (2012). *Theory of Society, Volume 1* (R. Barrett, Trans.; 1st ed.). Stanford University Press.

Lum, K., Swarup, S., Eubank, S., & Hawdon, J. (2014). The Contagious Nature of Imprisonment: An Agent-Based Model to Explain Racial Disparities in Incarceration Rates. *Journal of The Royal Society Interface*, *11*(98), 20140409. https://doi.org/10.1098/rsif.2014.0409

Maoz, Z., & Abdolali, N. (1989). Regime Types and International Conflict, 1816-1976. *Journal of Conflict Resolution*, *33*(1), 3–35. https://doi.org/10.1177/0022002789033001001

Maoz, Z., & Russett, B. (1993). Normative and Structural Causes of Democratic Peace, 1946–1986. *American Political Science Review*, *87*(3), 624–638. https://doi.org/10.2307/2938740

Mayntz, R., & Scharpf, F. W. (Eds.). (1995). *Gesellschaftliche Selbstregelung und politische Steuerung*. Campus.

McIntyre, L. C. (1993). Complexity and Social Scientific Laws. *Synthese*, *97*(2), 209–227. https://doi.org/10.1007/BF01064115

Mesoudi, A. (2011). *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. University of Chicago Press.

Mesoudi, A., Veldhuis, D., & Foley, R. A. (2010). Why Aren't the Social Sciences Darwinian? *Journal of Evolutionary Psychology*, *8*(2), 93–104. https://doi.org/10.1556/JEP.8.2010.2.1

Mesoudi, A., Whiten, A., & Laland, K. N. (2006). Towards a Unified Science of Cultural Evolution. *Behavioral and Brain Sciences*, *29*(4), 329–347. https://doi.org/10.1017/S0140525X06009083

Oppenheim, P., & Putnam, H. (1958). Unity of Science as a Working Hypothesis. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota Studies in the Philosophy of Science. Vol. 2* (pp. 3–36). University of Minnesota Press.

Parsons, T. (1937). *The Structure of Social Action: A Study in Social Theory with Special Reference to a Group of Recent European Writers*. McGraw Hill.

Parsons, T., & Shils, E. A. (Eds.). (1951). *Toward a general theory of action* (pp. xi, 506). Harvard University Press. https://doi.org/10.4159/harvard.9780674863507

Potochnik, A. (2017). *Idealization and the aims of science*. The University of Chicago Press.

Reydon. (2021). Generalized Darwinism as Modest Unification. *American Philosophical Quarterly*, *58*(1), 79. https://doi.org/10.2307/48600687

Reydon, T. A. C., & Scholz, M. (2009). Why Organizational Ecology Is Not a Darwinian Research Program. *Philosophy of the Social Sciences*, *39*(3), 408–439. https://doi.org/10.1177/0048393108325331

Richerson, P. J., & Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press.

Rosenberg, A. (2012). *Philosophy of Social Science* (4th ed.). Westview Press.

Rosenberg, A. (2017). Why Social Science is Biological Science. *Journal for General Philosophy of Science*, *48*, 341–369. https://doi.org/10.1007/s10838-017-9365-0

Schatzki, T. R. (2001). On Sociocultural Evolution by Social Selection. *Journal for the Theory of Social Behaviour*, *31*(4), 341–364. https://doi.org/10.1111/1468-5914.00164

Schimank, U. (2012). Markenbildung und Markenbindung auf dem Theorie-Markt-Eine Notiz zur Soziologie der Soziologie. *Zeitschrift Für Theoretische Soziologie*, *1*(1), 10–16.

Scriven, M. (1994). A Possible Distinction Between Traditional Scientific Disciplines and the Study of Human Behavior. In M. Martin & L. C. McIntyre (Eds.), *Readings in the Philosophy of Social Science* (pp. 71–78). MIT Press.

Sewell, W. H. Jr. (2005). *Logics of History: Social Theory and Social Transformation*. University of Chicago Press.

Smith, C. (2016). The Conceptual Incoherence of "Culture" in American Sociology. *The American Sociologist*, *47*(4), 388–415. https://doi.org/10.1007/s12108-016-9308-y

Smith, R. C. (2014). Black Mexicans, Conjunctural Ethnicity, and Operating Identities: Long-Term Ethnographic Analysis. *American Sociological Review*, *79*(3), 517–548. https://doi.org/10.1177/0003122414529585

Sober, E. (1992). Models of Cultural Evolution. In P. Griffiths (Ed.), *Trees of Life* (pp. 17–39). Springer. https://doi.org/10.1007/978-94-015-8038-0_2

Tanney, J. (2013). Causes versus Reasons in Action Explanation. In B. Kaldis (Ed.), *Encyclopedia of Philosophy and the Social Sciences* (pp. 73–78). SAGE.

Taylor, C. (1971). Interpretation and the Sciences of Man. *The Review of Metaphysics*, *25*, 3–51.

Tooby, J., & Cosmides, L. (1989). Evolutionary Psychology and the Generation of Culture, Part I: Theoretical Considerations. *Ethology and Sociobiology*, *10*(1–3), 29–49. https://doi.org/10.1016/0162-3095(89)90012-5 van Bouwel, J., & Weber, E. (2008a). A Pragmatist Defense of Non-Relativistic Explanatory Pluralism in History and Social Science. *History and Theory*, *47*(2), 168–182. https://doi.org/10.1111/j.1468-2303.2008.00445.x

van Bouwel, J., & Weber, E. (2008b). De-Ontologizing the Debate on Social Explanations: A Pragmatic Approach Based on Epistemic Interests. *Human Studies*, *31*(4), 423–442. https://doi.org/10.1007/s10746-008-9102-0

Weber, M. (1978[1921/22]). *Economy and Society: An Outline of Interpretive Sociology* (G. Roth & C. Wittich, Eds.; 4th ed.). University of California Press.

Wilson, E. O. (2000[1975]). *Sociobiology—The New Synthesis*. Belknap Press.

Winterhalder, B., & Smith, E. A. (2000). Analyzing Adaptive Strategies: Human Behavioral Ecology at Twenty-Five. *Evolutionary Anthropology: Issues, News, and Reviews*, *9*(2), 51–72.

Chapter 16

Equations at an Exhibition: On the Cultural Price Equation

Tim Lewens[i]

Abstract

The Price Equation is a powerful, and unusual, tool within evolutionary theory. Because it is completely general in application, and also very nearly free of distorting idealisations, the Price Equation is widely regarded as having exceptional power for understanding evolutionary change. It is no surprise, then, that it has been applied to many different contexts outside of traditional 'organic' evolution, including the domain of cultural evolution. In this essay I argue for various ways in which the Price Equation can mislead about cultural evolutionary theory. They all derive from difficulties that processes of cultural reproduction pose for attempts to distinguish 'selection' from 'transmission'. This does not mean the cultural Price Equation is of no use: its value remains as an analytical tool in those circumstances where a distinction between selection and transmission can be drawn without too much distortion.

Keywords: Cultural Attraction, Cultural Selection, Musorgsky, Price Equation

---

[i] University of Cambridge, Department of History and Philosophy of Science, Free School Lane, Cambridge CB2 3RH, Email tml1000@cam.ac.uk.

16.1 The Power of Price

It is often useful to reach an understanding of evolution in natural populations by assuming, for the purposes of analysis, that they are far simpler than is really the case. For example, one might assume that selection is the only force at work, or that organisms are asexual, or that mating is perfectly random, or that offspring traits correspond precisely to their parents' average values for those traits. The Price Equation is an unusual tool within evolutionary theory because it does not work like this (Birch 2013: 15). Instead, it captures relationships between average trait values in parental and offspring generations, which hold in populations even when affected by many complicating aspects. According to a common interpretation, the Price Equation can be used to attributes changes in average trait values across generations to different causal factors, in particular the influences of selection and transmission bias.

Because it is completely general in application, and also completely accurate with respect to the mathematical description it offers, the Price Equation is widely regarded as having exceptional power for understanding evolutionary change. It is no surprise, then, that it has been applied to many different contexts outside of traditional 'organic' evolution, including the domain of cultural evolution (e.g. Henrich 2004, Lehmann & Feldman 2008, Kerr & Godfrey-Smith 2009, Helanterä & Uller 2010, 2020, El Mouden et al 2014, Birch 2017, Nettle 2020). There is now considerable—and understandable—enthusiasm for this approach, which has perhaps reached its strongest expression to date in Baravalle and Luque's assertion that, 'a certain version of the Price equation is the fundamental law of cultural evolutionary theory' (2021: 1).

In this essay I argue for various ways in which the Price Equation can mislead about cultural evolutionary theory. They all derive from difficulties that processes of cultural reproduction pose for attempts to distinguish 'selection' from 'transmission'. This does not mean the cultural Price Equation is of no use, and it is not my intention here to undermine specific deployments of the Price approach in contexts such as debate over the evolution of altruism. The Price Equation remains a useful analytical tool in those circumstances where a distinction between selection and transmission can be drawn without too much distortion. But sometimes that distinction is distorting, and for that reason strong claims made for the power of the Price approach as a ‘heuristic device for discovering the actual causes of cultural change and accumulation’ should be rejected (ibid: 1).

16.2 A Brief Primer on the Price Equation

George Price (1970, 1995) thought his mathematical approach to selection could be used to capture *all* selection processes, whether they occurred in nature, the marketplace, or elsewhere. Even so, his approach has primarily been used to understand natural selection acting on populations of organisms, and it is useful for what follows to give a brief and intuitive presentation of how Price’s approach works in that context. Part of its power lies in its ability to capture selection acting at multiple levels (Okasha 2006, Price 1972), but for the purposes of this exposition very simple, single-level forms of selection will suffice.

The basic idea behind the Price Equation is intuitive. (The following outline draws on Okasha 2006, Frank 2005 and Price 1970, 1972.) Consider two successive generations of a population. There is no migration into, or out of, the population, so that all individuals in the later generation have parents in the earlier generation. Some individuals contribute more offspring than others. These facts about reproductive output may, or may not, covary with facts about other traits. If faster-running parents tend to contribute more offspring, and slower-running parents contribute fewer offspring, then, intuitively, there is selection on running speed. If, on the other hand, there is a completely random relationship between parental traits and reproductive output, then, intuitively, there is no selection at work. This association between a trait of interest—in this case, running speed—and reproductive output can be expressed by the statistical measure of covariance. Price’s idea is that the intensity of selection is captured by the expression:

$$cov(w, z)$$

Where w is the fitness of the members of the parental generation understood in terms of reproductive output, and z is some trait of interest—in this case, running speed. In a formal presentation it would be important to take into account a series of nuances regarding the precise notion of fitness in play, the question of whether generations are discrete or overlapping—i.e. whether some members of the parental generation persist into the offspring generation or not—and so forth. But these fine-grained issues are not relevant to the considerations of this essay.

If we knew that parents and offspring always resembled each other perfectly, or (for sexually reproducing organisms) that offspring's traits were always precise averages of parental traits, then the covariation term would be enough to calculate the change in average speed from one generation to the next. (Again, this should be fairly intuitive: if we know the degree to which fast runners have more offspring than slow runners, then we can predict not only that running speed will increase, but exactly how much it will increase by.) But usually in the biological world this resemblance is not perfect: developmental processes can make it the case that offspring and parents differ considerably. So, if one wants to keep track of how the average speed of the population changes, one also needs to know how much each parent's offspring differ from it on average. The overall modification to the selection-based prediction for population change must average all of these parent-wise changes, in a way that is weighted by parental fitness $E_w(\Delta z)$. (This fitness-weighting is required because, for example, one parent may have far more offspring than all the others: its degree of non-resemblance with respect to its offspring therefore needs to be given greater weight in calculating the overall population change.)

This now brings us to a standard simple formulation of the basic Price equation for a single level of selection. Afficionados will note that this particular version of the equation requires that I now use a term ω, representing *relative fitness*. This is simply the absolute individual fitness w from the expression above, divided by the population mean fitness $\overline{w}$. The change in the average value of a trait from one generation to the next $\Delta\bar{z}$ can be expressed as follows:

$$\Delta\bar{z} = cov(\omega, z) + E_w(\Delta z)$$

In other words, the Price equation gives us a way of understanding the overall change in the population's average value of some trait of interest. It is widely thought that the equation allows us to apportion change as due in some part to selection—understood as covariance between parental fitness and the trait in question—combined with some level of transmission bias—understood in terms of a fitness-weighted difference between values of the trait for parents and their offspring. If fast running is randomly related to fitness, but offspring are always faster than their parents (perhaps because of training, or nutrition) then the Price equation delivers the result that the change in average running speed is due entirely to transmission bias, and not to selection at all. Conversely, if offspring resemble their parents perfectly with respect to running speed, then the transmission term is set to zero, and any change in the population average will be attributed wholly to selection.

16.3Price on General Selection

The first extended effort to apply the Price Equation to the domain of culture comes from George Price himself (Price 1995). It appears in his discussion of a general approach to selection in a manuscript that was not published until after his death in 1995, but which is dated at around 1971. In his original 1970 paper on selection and covariance, Price had already briefly noted that his mathematical approach could encompass non-genetical selection (Price 1970). In that paper, he imagined a change in average IQ that might occur between a group of students who enrol on a given course, and those who successfully complete it (after others fail, or drop out). His covariance approach allows one to quantify the degree to which the course selects on IQ. In the later paper, where this general approach to selection is developed in much more detail, he uses an example that is, in many ways, peculiar. Price says that one might use his approach to capture the idea that the composer Musorgsky made a selection of certain properties when writing his musical suite *Pictures at an Exhibition*. Price explains, further, that one might quantify the intensity of selection on those properties (Price 1995).

In a moment I will say considerably more about this specific example, but one cannot understand Price's interest in it without explaining his understanding of what it would mean to have a fully general theory of selection. He was looking for an account that could represent *anything* that might be thought of as a 'selection'. Because of that, he gives some surprisingly banal examples. They are intended to demonstrate just how general the recommended approach is intended to be.

For example, one might be tempted to suppose that any evolutionary account of change must be focused on entities standing in reproductive relations to each other. This is encouraged by a standard practice (which I also followed at the beginning of this essay) of discussing the Price equation itself in terms of a 'parent' generation and an 'offspring' generation, with parents responsible for the production of differing offspring numbers. But Price thought of this as just one way in which his approach might work. He also wanted to be able to formalise the run-of-the-mill idea that an individual in a greengrocer's selects apples from a larger batch, keeping some and throwing others away. Here there is no literal 'parental' generation: the apples pre-selection do not reproduce to give rise to the apples post-selection. Instead, some apples persist and make it into the post-selection set, and others disappear. When this happens, one may find that the average quality in the population has changed.

Price has a particular way of conceptualising this process, which fits in with his mathematical approach introduced in section two above. One can think of some individual apples pre-selection as contributing, and others failing to contribute, to the post-selection population. Selection on quality occurs when an apple's pre-selection quality is non-randomly related to its contribution to the post-selection population; or, put another way, selection occurs when there is a non-random relationship between an apple's quality and its persistence. In other words, it is once again possible to represent the intensity of selection using a term $cov(w, z)$, where w now reflects the apples' persistence (rather than reproductive fitness), and z represents ripeness (or indeed any other trait that one might suspect is being selected, such as colour).

This means that Price sees selection processes everywhere: indeed, they are present in all cases where, as he puts it, 'packages' of some quantity of interest (e.g. individual apples, understood as packages of ripeness or colour) give rise to packages at a later time phase. Hence Price also points out that a series of flasks, with different concentrations of, say, saline solution, give rise to a 'selection' if different amounts are poured into beakers. The effect is that some contribute more than others (because lots of liquid is poured from some flasks, maybe none from others) to the beaker-bound 'packages' in the later population. If more is poured from the flasks with high saline concentration, and very little from the flasks with low saline concentration, then the average concentration across all packages will increase. Here ‘selection’ has occurred on saline concentration.

One of the reasons why Price’s equation is useful is because of the simple fact that if one knows, for example, for each apple whether it contributes or not, and one knows its pre-selection quality, one can compute the post-selection makeup of the population. Likewise, if one knows the concentration of saline in each flask, and how much it contributes to subsequent beakers, then one can also calculate the post-selection makeup of the population.

Note that Price’s initial examples all involve *stable* entities. It would not be possible to calculate the post-selection makeup of a population of apples from the facts of persistence and the quality of apples pre-selection if, for example, apples were the sorts of entities that had a tendency to change their quality (for better or worse) very quickly. This is why, in the case of organic evolution and many other processes involving changeable entities, it is also important to include a term in the Price Equation that accounts for transmission bias. As I mentioned above, if one knows, for example, that some organisms tend to contribute more to the next generation (in terms of offspring) than others, and that those making a greater contribution also tend to be fast runners, then one *cannot* conclude as a matter of mathematical certainty that average running speed will increase from one generation to the next. If fast runners and slow runners alike all tend to have slow-running offspring, then the population will change so that average running speed decreases.

16.4Pictures at an Exhibition

Now let me explain Price’s reference to Musorgsky. (Price himself uses the spelling ‘Moussorgsky’, but I will standardise throughout.) In doing so, I will give considerably more background than the 18-line paragraph (and accompanying diagram) that Price devotes to the example.

Pictures at an Exhibition is a series of fifteen short pieces of music. Victor Hartman—an architect and artist, as well as a friend of Musorgsky—died in 1873. In February 1874, a memorial exhibition was held for him. Vladimir Stasov, a member of Musorgsky's circle who worked in the art department of the St Petersburg Public Library, and who was one of the co-organisers of the exhibition, wrote that:

> One-half of these drawings shows nothing typical of an architect. They are all lively elegant sketches by a genre-painter, the majority depicting scenes, characters and figures out of everyday life, captured in the middle of everything going on around them… (Stasov, as quoted in Ross 1992: 16)

Stasov explained that:

> Musorgsky, who loved Hartman passionately and was deeply moved by his death, planned to "draw in music" the best pictures of his deceased friend, representing himself as he strolled through the exhibition, joyfully or sadly recalling the highly talented deceased artist (ibid.).

Hence the musical suite has a structure of a repeating 'Promenade', corresponding to the musician's walk around the exhibition, interspersed among ten pieces corresponding (although not without complications) to specific pictures from the exhibition.

Only eleven of the four-hundred works exhibited were chosen by Musorgsky as the basis for transformation into musical pieces. (In case this sounds confusing, note that two of the paintings were jointly represented in one piece. Hence eleven works give rise to ten pieces.) Price thinks of this as a selection by Musorgsky, on the reasonable grounds that while many (in fact, almost all) of the pictures make no contribution to Musorgsky's musical work, others do. One might quantify that contribution, either by assigning a value of one or zero according to whether a painting gives rise to a corresponding piece or not, or by assigning a number that represents how many bars of music are present in a picture's corresponding piece (which, again, will usually be zero). Recall that an apple either contributes (i.e. it is retained) or not (i.e. it is rejected) based on its quality. A flask contributes to some degree (i.e. its contents are transferred or not to a beaker) according to the concentration of solution in it. Likewise, says Price, a picture contributes a number of bars to a musical piece according to its instantiation of some quality.

Price illustrates the common framework for selection processes with this diagram (I accompany it with Price's own caption):

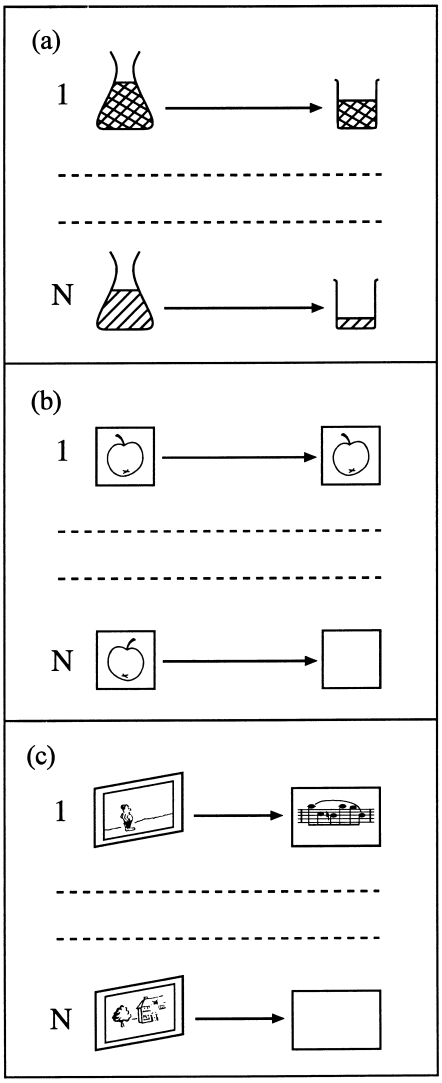


Figure 1: Three selection examples arranged in the pattern of the general selection model*. (a) The essential elements of the Fig. 2 example. (b) How the Fig. 1(a) example is fitted to the general model. (c) Musorgsky's selection of "Pictures from an Exhibition".*

Price uses the example of Musorgsky to illustrate some general features of his account that he thinks of as attractive. First, so long as pieces of music can be assigned to corresponding pictures—even if usually one makes use of the device that one picture corresponds to no music, while just a small number of pictures gives rise to several bars—then it is possible to make use of his framework, even though pieces of music and pictures are entirely different sorts of things. Second, Price's framework allows us to quantify the intensity of selection just so long as pictures yield musical bars as a non-random function of some property instantiated by the pictures. It does not have to be the case that music also instantiates that property.

It is worth lingering on Price's second point, because standard versions of the Price Equation *do* require that the property of interest can be measured in both pre- and post-selection generations. Okasha (2006: 24), for example, follows Rice (2004) in noting that, 'parental and offspring entities do not even have to be of the same type, so long as the character z is measurable on both.' This requirement follows from the fact that the investigator is usually interested in how some property changes in the pre- and post-selection populations, and also from the need to represent the effect of transmission bias. The transmission bias term relies on recording differences between parents and offspring with respect to the focal property.

Nonetheless, while the complete Price Equation does require that a property can be measured in both parent and offspring generations, Price's conceptual approach to the specific phenomenon of selection not require this. Remember that Musorgsky only composed pieces for 11 of the 400 pictures. Price suggests that:

> …if one could define interesting attributes of mood and subject matter that could be quantitatively evaluated in the paintings, one could measure (using definitions of 'selection intensity' given in Price 1972) the degree to which Musorgsky selected for or against these. (Price 1995: 394)

Perhaps one might try to quantify the 'sadness' of the paintings, how emblematic of 'Russianness' they are, or some measure of skill in their composition. One could then see how these properties in the pictures covary with the number of bars of music produced. Price does not require that it must be possible to measure those same properties of mood, or subject matter, in the pieces of music they give rise to. More generally, Price's approach could be applied to any property of the pictures, including their positioning in the exhibition room itself. Suppose, for example, I suspect that Musorgsky, a large and infirm man, wrote pieces for the eleven pictures that were the shortest walk from a particular spot in the original exhibition space, and ignored all the rest. (Alas, I could never really confirm this, because so little is known about the original exhibition.) Price's approach could quantify this selection on position of the original pictures. However, since Musorgsky's compositions are (i) musical pieces rather than physical pictures and (ii) entities that did not even exist until after the exhibition closed, it makes no sense to attribute qualities like 'position in the exhibition room' to the compositions themselves.

For Price, it makes sense to understand and quantify selection in a manner that is entirely independent of how one understands and quantifies transmission bias; so much so that his approach to selection can be applied in cases where his approach to transmission bias cannot. The significance of this point will become clearer when recent invocations of the cultural Price Equation are examined later in this chapter.

16.5Problems at an Exhibition

Price's specific example in some ways constitutes a best case for his approach. There really is a selection of sorts here, because while Musorgsky might conceivably have written music based on any one of the 400 pictures shown, in fact he only used 11. And so, one can attempt to analyse the degree of covariance between properties of interest in those 400 pictures, and the degree to which music arises from them.

That said, often the results of this kind of analysis will be silly. One could consider the four seasons of the year as the pre-selection set, and ask whether some contribute more minutes of music than others to Antonio Vivaldi's set of concerti *The Four Seasons*, understood as the post-selection set. In this case, all seasons contribute something. (This is unlike the Musorgsky case, where only a few paintings are translated into pictures.)

To avoid mixing up different kinds of things, I will refer to the meteorological seasons using English terms ‘spring’, ‘summer’ etc, and Vivaldi’s corresponding concerti using their Italian titles. *Wikipedia* claims that the usual performance times for La primavera and L’inverno (which correspond to colder meteorological seasons) are both a little shorter than L’estate and L’autunno (which correspond to warmer seasons, for in Italy autumn is warmer than spring). Supposing *Wikipedia* to be correct, and applying Price’s approach, one will conclude that there is selection on average monthly temperature at work, with warmer seasons giving rise to more minutes of music than colder seasons.

This verdict is obviously misleading. That is not by itself a problem for Price’s framework. He notes that an investigator must choose the property *x* to evaluate (that is, the property whose intensity will be assessed for covariation with productivity of the items in question) (Price 1995). In the case of *The Four Seasons* I decide, absurdly, to see whether average monthly temperature enjoys such covariation with production of minutes, and I find it does.

Price is aware that this may represent no genuine causal relation:

> If gold and silver are associated in some mineral, with the gold being easily apparent from its colour but the silver being difficult to recognize, then the gold content would be a primary selection property for a prospector who selected samples on the basis of appearance, while the silver content would be an associated property that would be indirectly influenced through the direct selection on a gold. (Price 1995: 395)

There are, however, deeper problems. Price is keen to stress that all genuine cases of selection can be captured by his approach. It is also important to ask whether all instances that his approach counts as ‘selection’ are genuine. The problem with a mapping from seasons to concerti is not that there is merely selection *of* average monthly temperature, but selection *for* some other property of the seasons that is correlated with monthly temperature. Rather, it seems misleading to think there is *any* genuine ‘selection’ going on here at all. True, the warmer meteorological seasons of summer and autumn can be linked to musical pieces that are longer than those linked to the colder seasons of winter and spring; but there is no reason to think this represents a causal contribution of *any* property of the seasons themselves, which gives rise to more minutes of music. The longer lengths of L’estate and L’autunno reflect nothing more than musical choices internal to the composition of the pieces, and entirely disconnected from the seasons themselves. In other words, if one tries to use Price’s formal approach all by itself to tell us when ‘selection’ occurs, then the results it delivers will sometimes be spurious. Instead, as Queller (2011: 10793-4) also suggests in a different context, Price’s approach needs to be deployed in circumstances where one already has reason to think one set of properties is causally influencing another.

There are other complications for Price’s approach that arise in the context of *Pictures at an Exhibition* itself. The repeating ‘Promenade’ theme—which, remember, is supposed to depict Musorgsky as he walks around the exhibition—is initially set apart in a clear way from the pieces corresponding to pictures. As the suite moves on, the promenade becomes integrated into some of the pictures themselves:

> Thus the Promenade theme, which represents Musorgsky himself, shows how by the end he has been drawn into the pictures and is no longer viewing them from the outside. (Russ 1992: 34)

This means that some of these later pieces do not merely represent, in musical form, a single picture: they also reflect Musorgsky’s experience of the whole exhibition. The piece ‘Con mortuis in lingua mortua’ cannot, in truth, be understood as corresponding solely to Hartmann’s painting of the Parisian catacombs. To some degree it should also be assigned to all the other 400 pictures, at least to the extent that those pictures affect Musorgsky’s mood, which in turn is woven into the piece. This is not merely an epistemic problem: it is not clear what facts (whether one can discover them or not) make it the case that, let’s say, some large fraction of the musical piece should be understood as corresponding to the catacombs painting, while other much smaller fractions are assigned to other paintings. One might try to decide this according to how far elements of the piece ‘resemble’ in mood one of the paintings or not. This would, however, be a mistake. I argued earlier that Price’s approach to quantifying the intensity of selection is supposed to proceed even when elements in the post-selection set do not possess the properties instantiated by the pre-selection set, hence even when the question of resemblance does not arise.

16.6The Price Equation in Cultural Evolutionary Theory

I anticipate that many readers will wonder why I raise such elaborate worries for Price’s own brief remarks on general selection. One might suspect they have no relevance for more developed efforts to apply Price’s approach in the context of mature forms of cultural evolutionary theory. In what follows, I will suggest that some of these recent approaches are, in fact, more prone to these conceptual difficulties than are Price’s initial suggestions for how his formal apparatus might be applied outside of organic evolution.

Before looking at detailed suggestions for how to apply Price’s approach to culture, it is worth reflecting on why it is often straightforward to apply that approach in the organic realm. The expression $cov(w, z)$ can be used to represent selection because, in this realm, it is usually clear enough which facts make it the case that an individual has some number of offspring. (I do not mean that it’s always easy to find out how many offspring an organism has, or that it’s always easy to find out what affects an individual’s production of offspring; only that it’s usually clear what constitutes an individual’s having some number of offspring.) Moreover, these facts are entirely independent of any question of how much an individual *influences* the offspring’s possession of whatever trait z one might be interested in. Indeed, this is part of the reason why a further term that corresponds to transmission bias is necessary. In other words, it is possible to distinguish neatly between (i) how productive an individual is, measured in terms of offspring number and (ii) the extent to which offspring properties resemble to parental properties.

Some ways of applying the Price Equation in the context of cultural evolution are similarly straightforward. That is because they focus on the extent to which what an individual has learnt covaries with how many *biological* offspring it has, and on the extent to which these biological offspring also acquire those same traits by learning. This is what Jonathan Birch (2017: 197) calls type-1 cultural selection, or CS_1. The approach to an entity's productivity is the same here as that used in the standard biological context, hence it can also be applied independently of questions about resemblance between parents and offspring.

Price's general approach to selection aims to maintain this key distinction between productivity and resemblance, albeit in an enlarged way. One can distinguish, for example (i) whether an apple persists or is discarded from (ii) whether the apple rots or ripens. One can distinguish between (i) how much a flask contributes to a given beaker and (ii) whether the solution contributed is stable in terms of its concentration. And one can distinguish (or at least it seems one can) between (i) how many bars of music a painting yields (which may depend on important features of the painting in question) and (ii) whether the music in question is even capable of instantiating important properties of the painting.

Birch's CS_1 is only one approach to understanding cultural selection. In other cases, instead of focusing on how learned traits affect an individual's production of biological offspring, investigators aim to understand an individual's production of 'cultural' offspring. Birch calls these latter approaches type-2 cultural selection, or CS_2 (2017: 199). The underlying motivation for CS_2 can seem intuitive enough: even if they have no biological offspring, some individuals nonetheless have plenty of cultural offspring, in the sense that their influence on the cultural traits of subsequent generations is significant. But remember, once again, that the sort of approach presupposed by the Price equation demands that two questions can be distinguished (Okasha and Otsuka 2020): (i) how many offspring (cultural or otherwise) does some entity have, and (ii) to what extent do the offspring resemble the parents? What I want to suggest here is that this distinction is far less clear-cut than one might think in the domain of culture, especially in the context of CS_2 approaches, and that this is the case because of well-understood features of cultural 'reproduction' (see especially Sperber 1996, 2000; Lewens 2015).

The analyst goes wrong if they end up determining how many cultural offspring an entity has by examining how many resembling cultural tokens there are in the next cultural generation.

That elides two sets of facts that are supposed to be distinct in the Price treatment. More importantly, that elision can mislead. Imagine, for example, that Shy Simon invents a wonderful gadget with features ABC. Meanwhile, Outgoing Oswald invents a mediocre gadget with features XYZ. No one pays attention to Shy Simon; lots of people pay attention to Outgoing Oswald. But in spite of paying lots of attention to Oswald, many folk in fact end up constructing a gadget just like Simon's, with features ABC. No one constructs anything with features XYZ.

Why might this sort of thing happen? It might naturally be explained using the resources of cultural attraction theory (Sperber 1996, 2000; Buskell 2017; Scott-Phillips et al 2018). Maybe a gadget with features ABC has the following properties: it answers a widely-experienced need, it makes use of elementary and intuitive design principles, it can be built easily using cheap and plentiful materials. It is, in Sperber's sense, an 'attractor'. Meanwhile, a gadget with features XYZ has all the opposite features: no one has much need for such an item, its mode of operation is not at all intuitive, its raw materials are expensive. Even so, it's not out of the question that individuals observe Outgoing Oswald, and they think 'I can do so much better than his XYZ!' And so, *because* they observe Oswald, they end up building gadgets with features ABC.

I do not think this kind of case is so far-fetched that one should dismiss it as irrelevant. But it does raise quite serious conceptual problems for some efforts to apply the Price approach to culture. First, it is not the case that the gadgets with features ABC are *offspring* of Shy Simon's gadget, even though they resemble it closely. Shy Simon's gadget plays no role in explaining the genesis of the resembling token gadgets. One could, therefore, try to argue in ways that parallel what might be said about organic reproduction above: an individual can have plenty of babies which, for whatever reason, fail to resemble their parents. So why not simply say that Oswald's gadget XYZ has plenty of cultural babies, which fail to resemble it? The Price approach would then give a peculiar result, namely that this is a case with intense selection of features XYZ (which no one is impressed by) accompanied by very strong transmission bias. This result highly peculiar and unilluminating. While Oswald's gadget—unlike Simon's—has a causal role in explaining the production of these later tokens, that role is small. There are many other factors that also explain (causally) the production of later tokens—everything from the availability of raw materials to the intuitive nature of the design. It is not clear why one should give Oswald's gadget special credit here in accounting for cultural 'fitness'.

16.7El Mouden et al on Cultural Selection

El Mouden et al (2013) have developed a prominent approach to using the Price Equation in the context of what Birch calls CS_2. Their formulation of the Price equation for culture mirrors the standard formulation for organic evolution exactly: indeed, the equation is written in the same way in both cases, albeit with the letter c (rather than ω) representing some measure of cultural 'fitness' (ibid: 233):

$$\Delta\bar{z} = cov(c, z) + E_c(\Delta z)$$

The intuitive idea behind this is, again, simple. Suppose a population is once again divided into parental and offspring 'generations'. These needn't be biological parents (any more than Price's pre-selection apples are parents of the post-selection apples), or indeed 'parents' in any standard sense of the term. Instead, each individual is thought of as productive of some number of cultural descendants: some individuals may have none, others many. Imagine that individuals with short hairstyles have more cultural descendants than individuals with long hairstyles. If one knows that cultural descendants resemble their cultural parents perfectly, then it is easy calculate that average hair length will go down in the subsequent cultural 'generation'. That said, this calculation will need to be modified if individuals' cultural descendants do not resemble them perfectly, or if they resemble them only poorly.

At first sight this approach might seem to work without too much trouble: perhaps some individuals in a population—think of them as social influencers—have lots of disciples, and those disciples aim to mirror the hairstyles of their heroes. Other individuals have next to no one paying them any attention, and no one makes much effort to mirror their hairstyles. One can then see how hair length changes from one cultural 'generation' to the next: cultural selection on hair length will be very strong if individuals with lots of disciples also have longer hair, while individuals with few disciples have shorter hair. Cultural selection might be overwhelmed if, for some reason, the efforts of disciples to mirror the hairstyles of their heroes frequently fail. Maybe they attempt to style their hair at home, and they end up cutting far too much off. Under these circumstances the term that represents ‘transmission bias’ will be high, because offspring tend not to resemble their cultural ‘parents’ very closely.

16.8Problems for the Cultural Price Equation

This may sound promising, and doubtless the Price approach is useful in some contexts. But care must be taken when handling the approach. First, some of El Mouden et al’s remarks about their framework do not really hit the mark. Recall, for example, that the transmission term in the cultural Price Equation is zero just so long as offspring resemble parents perfectly with respect to a trait of interest. This is so regardless of which causal mechanism might explain why cultural offspring and their parents are similar. A contrived example will illustrate the difficulties this causal agnosticism poses.

> **Bake Off!** Baking ‘influencers’, with many disciples, manage to invent a new baking technique DEF. DEF is responsible for producing exceptionally delicious cakes. Meanwhile, individuals with only a tiny handful of disciples invent a poor baking technique PQR. Use of PQR yields disgusting cakes. The different bakers’ disciples all manage to successfully re-produce the techniques of their mentors, and average cake quality increases in the population.

The cultural Price equation represents this as a case where transmission bias is low or non-existent, and cultural selection is wholly responsible for the population-level change. But suppose that while the influencers’ disciples are inspired by their heroes to make the most delicious cakes they can, they do not copy their heroes’ techniques: instead, they use their creativity and judgement. Because technique DEF is not too hard to discover, and objectively excellent—once again, it is an ‘attractor’—the disciples of the more influential bakers manage to arrive at it independently.

This example shows that nuance must be added to El Mouden et al's claim that, '…an important part of cultural evolution is the transmission component—which reflects the action of minds that have been shaped by natural selection to process information in ways that enhance genetic fitness' (2013: 237). The transmission component may indeed reflect the action of minds in this way, but the selection component—in Price's sense of 'selection'—can also reflect the action of minds in the very same way. That is because the Price approach understands 'selection' to have occurred merely when offspring traits resemble parental traits, thereby making the transmission bias term small. The creative use of individuals' minds can potentially bring about both divergence and convergence when those individuals approach the same problem as their cultural 'parents'. Divergence is likely when there are many viable solutions, the problem itself is only vaguely specified, and there are few constraints on the approaches likely to be followed. Convergence is more likely when there are only a few viable solutions, the problem itself is tightly specified, and approaches likely to be followed are highly constrained.

Baravalle and Luque offer the suggestion that:

> …if an evolutionary factor is able to modify the 'content' of a cultural variant (i.e., its characteristic features, or the kind of behaviour that is related to it), then it should be classified as a transmission bias; if an evolutionary factor only modifies the distribution of a cultural variant, then it should be taken as a selective–or, more generally, 'sorting'–force. (2021: 5)

Note that this approach only works if by 'modify the content' one means 'modify the content *in such a way that it changes from the parental generation*'. Suppose, for example, individuals begin with vaguely formed ideas that are quite different to those of their cultural 'parents'; but various processes of reasoning refine those vague ideas, and result in eventual convergence across cultural 'generations'. The content of the offspring generation's views has been modified here, but in a way that results in the diminution of transmission bias, because that modification reduces dissimilarities between parents and offspring.

El Mouden et al go on to say that, '…the disagreement between those who advocate a Darwinian or a non-Darwinian approach to cultural evolution comes down in large part to different views about the relative importance of selection versus transmission in cultural change' (2013: 238). Again, this isn't quite right: at least some disputes—for example, over the relative importance of cultural 'attraction'—do not concern how important the transmission term is. In the example used above, the better technique DEF is an attractor, and it is because of its status as an attractor that transmission bias is so low.

It is now easier to understand why some advocates of cultural attraction theory complain that it is misleading to equate 'attraction' with 'transmission bias' (as El Mouden et al seem to do). Some advocates of cultural 'attraction' instead point out that (as I have just explained) various forms of creative inference can potentially underpin cultural change whether the transmission term is significant or not (Scott-Philipps et al 2018). The worry here is not that selection might be less important than transmission; it is that the distinction between selection and transmission—which the Price Equation encourages the analyst to understand as distinct causal factors represented by the two terms on the right hand side of the equation—is a misleading one because it does not map neatly onto underlying cognitive processes. The very same factors of cultural 'attraction' that can underpin cultural selection in a context where some influential parents have traits that already correspond to 'attractors', can instead underpin cultural transmission bias when influential parents instead have traits that do not correspond to attractors.

16.9Cultural Reproduction

I mentioned above that Price's general approach to selection relies on a strict distinction between the facts that make it the case that a pre-selection entity is productive with respect to the post-selection set, and the facts that make it the case that elements of the post-selection set resemble elements of the pre-selection set. Recall, for example, that it is one thing to determine whether an apple persists or is eliminated, another to determine whether the apple changes its properties over time. In Price's example of Musorgsky, some of Hartmann's pictures are productive with respect to the generation of music, others are not; but these facts are supposed to be determinable independently of any question of whether the music resembles the pictures.

El Mouden et al abandon this strict distinction. Rather than keeping the questions of whether an individual has cultural offspring distinct from questions of resemblance, they are instead blurred. They begin by defining cultural ancestry as follows: ‘Person A is a cultural ancestor of person B if the value of z person B has was influenced by the value of z person A had’ (2013: 233). This definition of ancestry in terms of influence departs from Price’s treatment. The same may be true of the approach suggested by Kerr and Godfrey-Smith (2009: 533). They note that the Price Equation requires that lines of ‘connection’ can be drawn between parents and offspring. In the biological context these are usually reproductive relations. Kerr and Godfrey-Smith continue: ‘we have discussed connections mostly as parent–offspring relations…Alternatively, a connection may represent other forms of *influence* between entities such as material or information flow’ (emphasis added). But this notion of connection as ‘influence’ once again threatens to undermine the distinction that is central to Price’s approach, because an intuitive way to understand ‘influence’ is partly via the notion of resemblance of token entities across generations. I stressed in section four above that, for Price, a given picture can be an ancestor of a piece of music even if the music and the picture have no shared properties, hence no shared trait z whose value in the musical piece is influenced by the picture. Or consider a biological case where inheritance is exceptionally unreliable: fast runners have lots of offspring, slow runners have very few, but fast running parents fail to have fast running children. The investigator does not determine which organisms are an individual’s parents by asking which organisms have influence on the individual’s running speed: indeed, this is a central case of selection for running speed being strong even though individuals are not influenced by their parents at all in this respect.

El Mouden et al move on to suggest that, ‘cultural fitness is a measure of cultural influence, reflecting both the number of people who learn from an individual, and the degree to which their traits are influenced when they do learn’ (2013: 233). Again, this idea of equating cultural fitness with cultural influence *on traits of the offspring generation* goes against Price’s background conceptualisation of selection: for here, recall, a painting might have high cultural fitness merely because it gives rise to many bars of music. The question of whether traits of the music are ‘influenced’ by the picture does not come into it.

El Mouden et al say a little about how they understand this notion of cultural influence—specifically, they claim that an individual may be influenced by many different individuals, to different degrees, for a specific trait—but this does not go far to explaining what counts as strong or weak influence. Suppose, for example, that I am so disgusted by an individual's behaviour that I try my best to act in the exact opposite way: does the individual in question influence me strongly, because their effect on my life is marked, or weakly, because I do not aim to be anything like them? On the face of things, if an individual is highly influential with respect to future generations, one might assume this means that it is because of exposure to that individual, that those in subsequent generations act or look a certain way that approximates to that of the influencer. But note again that this defies what the Price equation asks us to do, which is to distinguish between (i) the productivity of an entity strictly with respect to how many offspring it generates, whether an entity persists or not, whether it gives rise to bars of music, and so forth and (ii) the degree to which the post-selection entities correspond to the pre-selection entities.

In a footnote to their recent paper on the cultural Price Equation, Baravalle and Luque (2021: 5) note this problem. They thank Jonathan Birch for drawing their attention to it in correspondence. Their response is very brief, and invokes Lewontin's distinction between dynamical and empirical sufficiency of a model. They understand this as a distinction between a model's ability in principle to capture the causal processes underlying some change, and the investigator's ability to measure and record those processes. Unfortunately, I do not see how this distinction helps in this type of case, where the Price approach distorts the nature of the underlying causal processes.

If 'influence' is supposed to correspond to an individual's cultural fitness—as opposed to some hybrid quality that corresponds in part to cultural fitness, and in part to some degree of success in transmission of a given trait—then it should be possible to conclude that an individual has been highly influential, in the sense of producing entities in the offspring generation, even though those entities do not resemble the parental individual at all. It is difficult, though, to see what that could possibly mean in practice, because what makes the analyst conclude that an individual has influence is precisely the extent to which they cause others to be similar to them. Again, this is not like the cases Price focuses on: in the biological case it is perfectly easy to see how one might conclude that an individual is extremely fit, in the sense of having many babies, without needing to assess the individual's 'influence' on those babies with respect to some trait or another. It is also easy to see how one might conclude that one apple does, and another does not, succeed in persisting into the post-selection set, again without having to ask anything about whether the traits of an apple post-selection are 'influenced' by traits pre-selection.

El Mouden et al's mathematical approach gives some clues about how their notion of influence is understood. As one might imagine, the application of the Price equation leads them to assert that the change in average trait value from one generation to the next is given by 'the population mean of the weighted sum of cultural influences on each individual, plus the population mean of the extent to which each descendant individual j spontaneously departs from his cultural influences' (2013 Supporting Document S1: 1-2). This means that if descendant individuals do *not* spontaneously depart from their cultural influences, any change in population mean *must* be attributed wholly to that weighted sum of cultural influences on each individual. In other words, cultural influence is understood as whatever accounts for population change in the mean value of a focal trait, once any transmission bias is set aside. This is not a desirable result, because it makes it irrelevant to ask whether the individuals in question are, or are not, also using their own cognitive apparatus in creative ways that explains why they resemble (rather than depart from) their cultural influencers. Another way of putting this is that 'cultural influence', for El Mouden et al, ends up including any spontaneous creative role that the members of offspring generation may exercise in bringing about cultural resemblance.

To see this more clearly, consider two different scenarios again related to baking. Assume for the sake of argument that in both cases individuals in the offspring group only encounter one individual each in the parental group. In this sense, they are only ever 'influenced' by one individual when it comes to determining their cultural phenotype for baking, because (by hypothesis) there is no one else for them to be 'influenced' by. Even so, some individuals in the parental group have many disciples, while others have just a few. Individuals with many disciples formulate a new baking technique GHI, while individuals with few disciples instead formulate technique RST. GHI results in much better cakes than RST. Suppose, once again, that the disciples of the bakers using GHI manage to arrive at technique GHI; meanwhile, the disciples of the bakers using RST also manage to formulate RST. Because the techniques of offspring perfectly resemble those of their cultural parents, it immediately follows that transmission bias is set at zero. Hence it also follows that *all* of the change in the population mean should be attributed to 'the population mean of the weighted sum of cultural influences on each individual' (hence to cultural selection), regardless of any further details one might learn about *why* offspring resemble parents.

There are, however, many different possibilities for why this resemblance might occur, and the Price Equation does not discriminate between them. This is a limitation of the approach. Here are just two such possibilities:

> **Variant one: slavish copying.** All the disciples set out with the goal of doing precisely what their heroes do. They study their heroes' YouTube videos, and copy their every movement slavishly.

> **Variant two: the drive to improve.** Disciples all admire their respective baking heroes, but they make no effort to copy their techniques in any respect. They want only to be better bakers. Even so, it turns out that disciples of the GHI bakers all end up formulating GHI, and disciples of the RST bakers all end up formulating RST. Why the resemblance? The majority of bakers in the offspring generation are highly talented at baking, a few are not so good. The more talented bakers—because of their good judgement—also pick their heroes well. They admire bakers who are talented enough to invent GHI, while poor bakers end up admiring those who instead are only good enough to invent RST. And because of this difference in skill, when the better bakers in the offspring generation set out to devise a technique, the result of their talents is that they also develop GHI, while the poorer bakers develop RST.

El Mouden et al's approach demands that both scenarios be analysed in exactly the same way. Because offspring resemble their cultural parents perfectly with respect to their baking techniques, the transmission bias term must be set at zero. The cultural Price Equation therefore mandates that all the population change is attributed to cultural influence from the parental population, which is synonymous with cultural selection.

On the face of things this is a shortcoming: it seems important to find some way of recognising important differences between the two scenarios. First, the influence of the parental population is lower in scenario two than scenario one, because in scenario two the offspring individuals' own creativity—rather than the influence of their heroes—takes more responsibility for the change in the direction of better cakes. Second, it seems a mistake to attribute *all* of the overall population change to selection in scenario two. What happens here is talented bakers (i) admire other talented bakers and also (ii) they are better at discovering valuable baking innovations; conversely, untalented bakers admire other untalented bakers, and they are worse at discovering valuable innovations. This looks like a 'false positive' for Price's approach: it meets the criteria for there being selection, but on inspection it is not clear why this really counts as a case of selection at all.

16.10 Diagnosis

I have raised a series of problematic features for efforts to apply a Price Equation approach to understanding cultural change. Is there anything of a systematic nature that can be said to explain why the Price approach can be troublesome in this domain? The answer, it seems to me, lies in an important way in which cultural 'reproduction' differs from organic 'reproduction': it is a difference that has been stressed most prominently by Sperber and followers (Scott-Phillips et al 2018; see also Nettle 2020 for some sceptical caveats).

Price's basic approach, as I have stressed, relies on a strong distinction at the level of underlying processes between (i) the question of how many elements in the post-selection set are produced by elements in the pre-selection set and (ii) the question of the extent to which elements in the post-selection set resemble elements in the pre-selection set. In the organic context this distinction is usually unproblematic. It is perfectly coherent to judge, for example, that organism A has three offspring which, in fact, are much more like organism B; meanwhile, organism B has one offspring that is just like organism A.

In some of Price's motivating examples of non-organic selection the same distinction is also unproblematic: it is perfectly coherent to judge that apple A persists, while rotting in such a way that it ends up with the qualities of texture and taste had by apple B; meanwhile, apple B is thrown away. The same distinction is far harder to apply in cultural contexts. For suppose an individual's token behaviour B disgusts many other individuals; they are shocked by that behaviour, and try to model something very different. Should one say that although token behaviour B has many cultural offspring, those offspring end up resembling a wholly different token behaviour A? If one does say this, absurdities follow in how we parse these situations—the whole point is that while observers are certainly affected by B, there is no helpful sense in which B is being selected, or is highly 'fit'. The individuals in question are deliberately avoiding any repetition of B.

The alternative is also unpalatable. One can assign a low cultural fitness to B on the grounds that items resembling B do not end up appearing reliably in the offspring generation. This leads to a different problem: one is no longer drawing the sort of distinction the Price approach mandates between selection and transmission bias. Instead, the assessment of B as of low fitness reflects an amalgam of both the amount of 'productivity' B has, in the sense of the number of people who are in some way affected by B, and also the fact that B's 'offspring' are usually nothing much like B. In other words, the Price approach either gives the wrong result, or it gives the right result at the expense of blurring the distinction between the two factors it aims to keep separate.

The more basic problem, which these scenarios draw attention to, is that there are no cultural 'gametes' transmitted from the first individual to these various others. If there were, then one could find in nature a strict division between questions of how many offspring a cultural individual has—i.e. how many sets of gametes it gets into the next generation—and questions of whether its offspring resemble it to a greater or lesser extent—i.e. whether those gametes develop in ways that mean offspring and parents are alike. Of course, in some of Price's other examples there are also no 'gametes'—but there are easily trackable entities (a persisting apple, a corresponding beaker into which a flask is poured) that again allow the analyst to distinguish in a neat way between how 'productive' an entity is, and how much its products resemble it.

Price's example of Musorgsky appears, on the face of things, to fit this paradigm well: one might think it is not difficult to distinguish between how much music a picture gives rise to, and how much the music in question resembles the corresponding picture. Indeed, the first question can often be answered even when the second one makes no sense. But, as I argued, there are circumstances where even for *Pictures at an Exhibition* these questions become hard to untangle. It is certainly not a general feature with application across all cultural change.

This does not mean that Price's approach should be rejected. I have indicated that Pricean approaches to what Birch calls CS_1 do not run into the problems posed in this essay. There is, however, a dilemma for the most enthusiastic proponents of a Pricean approach to cultural evolution. If one saves Price's approach by claiming that it only represents those domains of cultural evolution where one can draw clear distinctions between facts about productivity and facts about resemblance, then one will also have to conclude that it is adequate only for the representation of a subset of instances of cultural change.

Acknowledgements

I am extremely grateful to André Ariew, Jonathan Birch, Olesya Bondarenko, Andrew Buskell, Azita Chellappoo, Mitchell Distin, Daniel Nettle, Samir Okasha, Raphael Scholl and Philipp Spillman for comments on an earlier draft.

References

Baravalle, L. and V. Luque (2021) 'Towards a Pricean Foundation for Cultural Evolutionary Theory' *Theoria* doi: 10.1387/theoria.21940

Birch, J. (2013) *Kin Selection: A Philosophical Analysis*. PhD Dissertation, University of Cambridge.

Birch, J. (2017) *The Philosophy of Social Evolution*. Oxford: Oxford University Press.

Buskell, A. (2017) 'What are Cultural Attractors?' *Biology and Philosophy* 32: 377-394.

El Mouden, C., J.-B. André, O. Morin and D. Nettle (2013) 'Cultural Transmission and the Evolution of Human Behaviour: A General Approach Based on the Price Equation' *Journal of Evolutionary Biology* 27: 231-241.

Frank, S. (1995) 'George Price's Contributions to Evolutionary Genetics' *Journal of Theoretical Biology* 175: 373-388.

Helanterä, H. and T. Uller (2010) 'The Price Equation and Extended Inheritance' *Philosophy and Theory in Biology* 2 (201306): 1-17

Helanterä, H. and T. Uller (2020) 'Different Perspectives on Non-Genetic Inheritance Illustrate the Versatile Utility of the Price Equation in Evolutionary Biology' *Philosophical Transactions of the Royal Society B* 375**:** 20190366.

Henrich, J. (2004) 'Cultural Group Selection, Coevolutionary Processes and Large-Scale Cooperation' *Journal of Economic Behavior and Organization* 53: 3-35.

Kerr, B. and P. Godfrey-Smith (2009) 'Generalization of the Price Equation for Evolutionary Change' *Evolution* 63: 531-536.

Lehmann, L. & M. Feldman (2008) 'The Co-Evolution of Culturally Inherited Altruistic Helping and Cultural Transmission under Random Group Formation' *Theoretical Population Biology* 73: 506-516.

Lewens, T. (2015) *Cultural Evolution: Conceptual Challenges*. Oxford: Oxford University Press.

Nettle, D. (2020) 'Selection, Adaptation, Inheritance and Design in Human Culture: The View from the Price Equation' *Philosophical Transactions of the Royal Society B* 375: 20190358.

Okasha, S. (2006) *Evolution and the Levels of Selection*. Oxford: OUP.

Okasha, S. and J. Otsuka (2020) 'The Price Equation and the Causal Analysis of Evolutionary Change' *Philosophical Transactions of the Royal Society B* 375: 20190365.

Price, G. (1970) ‘Selection and Covariance’ *Nature* 227: 520-521.

Price, G. (1972) ‘Extension of Covariance Selection Mathematics’ *Annals of Human Genetics* 35: 485-490.

Price, G. (1995) ‘The Nature of Selection’ *Journal of Theoretical Biology* 175: 389-396.

Queller, D. (2011) ‘Expanded Social Fitness and Hamilton’s Rule for Kin, Kith, and Kind’ *PNAS* 108: 10792-10799.

Rice, S. (2004) *Evolutionary Theory*. Sunderland, MA: Sinauer.

Russ, M. (1992) *Musorgsky: Pictures at an Exhibition*. Cambridge: Cambridge University Press.

Scott-Phillips, T., S. Blancke and C. Heintz (2018) ‘Four Misunderstandings about Cultural Attraction’ *Evolutionary Anthropology* 27: 162-173.

Sperber, D. (1996) *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.

Sperber, D. (2000) ‘An Objection to the Memetic Approach to Culture’ in *Darwinizing Culture: The Status of Memetics as a Science*, Robert Aunger, ed., 163–74. Oxford: Oxford University Press.

Chapter 17

Unlike Agents: The Role of Correlation in Economics and Biology

Hannah Rubin[i]

Abstract

While there are many important similarities between evolution in biology and learning in economics, we should be cautious when importing ideas from one evolutionary context to the other. I will argue that there is a lack of caution is behind the tendency to think of measures of correlation (e.g., ‘relatedness’) as akin to attitudes of economic agents (e.g., as capturing how much an organism or agent ‘values’ or ‘cares about’ a social partner), leading to use of unreliable heuristics and misunderstandings in biology, as well as to misuse of biological results in economics.

---

[i] Department of Philosophy, University of Notre Dame, Notre-Dame, Indiana. Email: hannahmrubin@nd.edu.

17.1 Introduction

There are striking similarities between natural selection in biology and decision making in economics; both are, in a sense, optimizing processes. As Elliott Sober puts it: “Just as the (objectively) fittest trait evolves, so the (subjectively) best action gets performed” (Sober, 1998, p. 409). Borrowing of ideas and techniques between the two fields has been enormously beneficial and has led to mathematical frameworks that describe behavior in both fields. For instance, evolutionary game theory can describe evolution of social behaviors both in the biological context describing natural selection (Maynard Smith, 1982) and in the context of economics, sociology, anthropology, etc. describing (boundedly) rational agents learning and updating their strategies (Axelrod, 1984).

However, there are disanalogies between evolution and learning as well: things like genetic constraints generally do not make sense in the context of rational choice, and biological agents do not generally imitate their most successful neighbor, as economic agents often do. Ideally, we ought to outline the limitations of borrowing between fields so we can safely enjoy the benefits while avoiding potential pitfalls of being misled. While many limitations of the borrowing of ideas between economics and biology have been noted,[1] I will identify an overlooked type of limitation and outline the negative consequences of not paying attention to it.

Specifically, I will argue that there are pitfalls that arise from importing agential descriptions into biology in order to describe measures of ‘relatedness’ as a measure of ‘common interest’ between organisms or the degree to which one organism ‘cares’ about its social partner’s reproductive success. While this provides a helpful way to give intuitive explanations for why certain social behaviors are beneficial, there are limitations to the borrowing of the ideas like ‘common interest’ from the economic context. Below, I will briefly introduce the relevant biological concepts, then argue that overlooking these limitations has led to at least three problems in biology and economics: reliance on an unreliable heuristic calculation of inclusive

---

[1] For instance, because of how traits can be genetically encoded, there limitations to viewing ‘mother nature’ or selection itself as an economic agent choosing traits to climb peaks of an adaptive landscape (Okasha, 2018).

fitness, misuse of biological results in economics, and incorrect conclusions regarding the necessity of inclusive fitness for understanding the appearance of design.

17.2 Inclusive fitness, relatedness, and correlation

Inclusive fitness is seen as essential to explaining the evolution of *social behavior*, where how well an organism does, in terms of its reproductive success, depends both on the trait it has and the traits of its social partners. In the inclusive fitness framework, one looks at the effects an organism has on other organisms' reproductive success, taking into account its 'relatedness' to those organisms, rather than just looking at the organism's own reproductive success. The idea that relatedness between organisms can help explain social behaviors has been part of evolutionary theory since Darwin, but the theory of inclusive fitness introduced by Hamilton (1963, 1964) showed how precisely to take relatedness into account (see Dugatkin, 2007, for a historical overview). To get a picture of how the inclusive fitness framework allows us to view the consequences of social traits in a different way, we can look at how to calculate inclusive fitness and the related concept of neighbor-modulated fitness (which corresponds more closely to our standard notion of fitness as measuring an organism's reproductive success).

Roughly, the neighbor-modulated fitness of an organism is calculated by adding up the number of offspring the organism is expected to have. If we wanted to calculate the neighbor-modulated fitness of, for example, an altruist[2] and non-altruist, we would look at all the effects on the focal organism's fitness: whether it pays the cost c and whether it receives the benefit b. The neighbor-modulated fitness of an altruist is

$$-c + P(A_j|A_i) \cdot b$$

This captures the fact that an altruist always pays a cost and has some chance of receiving the benefit. This chance is given by $P(A_j|A_i)$, which is the probability that the organism's social

---

[2] For simplicity, we can think of an altruistic behavior as one where there is some cost to the focal organism (in terms of fitness – decreasing its survival probability or reproductive output) and some benefit to another. Of course, there are other sorts of behaviors we might care to explain, some of which will be discussed below.

partner (labeled organism j) is an altruist, given that the focal organism (organism i) is an altruist. The neighbor-modulated fitness of a non-altruist is

$$P(A_j|N_i) \cdot b$$

A non-altruist does not pay a cost, but still has some chance of receiving the benefit (which is the probability the social partner is an altruist, given that the focal organism is not). These calculations tell us the expected reproductive success of each type of organism, corresponding to our standard notion of fitness.

Inclusive fitness is an alternative mathematical framework in which fitness calculations track the offspring *caused by* a particular organism, rather than tracking the offspring an organism actually has. The offspring caused by the organism are then weighted according to a 'relatedness' parameter. What relatedness is will be discussed more just a moment, but for now we can note that it measures how likely it is that organisms will have the same trait (or the same genes). Then, the inclusive fitness of an altruist is

$$-c + Rb$$

An altruist affects its own fitness by $-c$ (it pays a cost) and its social partner's fitness by b (it provides a benefit, which we then weight by relatedness, R). The fitness of the non-altruistic trait is zero because it does not affect its own fitness or the fitness of its social partner in any way, relevant to our trait of interest.[3]

The inclusive fitness framework might initially seem counter-intuitive, so it is helpful to mention a basic observation: in general, a trait will increase in frequency when organisms with that trait have more offspring than the average organism in the population. Inclusive fitness gives us the information to determine whether a trait will increase in frequency by telling us how many offspring are caused by an organism and how likely it is that these offspring are had by an organism with the trait of interest.

It is important to emphasize, for the purposes of this paper, that relatedness is a measure correlation between types. Specifically, R measures how likely it is that the focal organism and

---

[3] Technically, these fitness calculations include a baseline non-social fitness component, which is omitted here because it is the same for both inclusive fitness and neighbor-modulated fitness.

its social partner share genetic material, relative to the rest of the population. More specifically, the relatedness of a focal organism (organism i) to its social partner (organism j) is:

$$R = P(A_j|A_i) - P(A_j|N_i)$$

or, the probability the social partner is an altruist given the focal organism is, minus the probability the social partner is an altruist given the focal organism is not.[4]

Though relatedness is a measure of correlation,[5] it is also often described as “a measure of the extent to which... the focal individual values its social partners...” (West and Gardner, 2013, p. R578). The idea behind this description is that if we are thinking about a focal organism wanting to pass on its genes, and relatedness is telling us how likely it is that the social partner has these same genes, we can think of relatedness as measuring how much a focal organism cares about its social partner’s fitness. This is supposed to be in contrast to neighbor-modulated fitness, where the probabilities of interacting with like individuals measures the extent to which “social partners have a similar disposition for altruism” (p. R578).

This interpretation of relatedness as how much a focal organism values its social partners is analogical, or a way of helping us understand how this term, which is a measure of correlation between types, could be used to explain the evolution of altruistic behaviors. However, as I will argue, this reliance on agential language has led to slippage and confusions regarding the transportability of concepts between economic and biological contexts. As an example of this, consider how Kevin Zollman describes the role relatedness plays in the evolution of honest communication in cases where there are conflicts of interest, e.g. between parents and offspring. In this case, honest communication can be seen as a type of altruistic action because it is costly for the honest organism, but beneficial for their relative. Zollman claims, however, that “the most popular solution to the biological altruism problem, inclusive fitness theory, cannot help in this context, since parent–offspring conflicts arise despite the high relatedness between

---

[4] For a discussion of when this definition of relatedness is equivalent to other common definitions of relatedness derived from the Price equation, see Rubin (2018).

[5] This is something that is agreed upon by inclusive fitness theorists. See, e.g., Marshall (2015), and references therein.

parents and offspring" (2013, p. 130). Instead, since it is well-known that correlations between traits can allow altruism to evolve, he proposes that we look to solutions using correlation and notes that "Relatedness might, beyond inclusive fitness, introduce additional correlation" (p. 131).

Zollman is not alone in contrasting relatedness and correlation in this way; in fact, as I will argue, similar tendencies to construe relatedness as akin to an attitude of an economic agent have caused problems in both biology and economics. First, I will argue that taking this interpretation of relatedness too seriously is a big reason for the reliance on heuristic methods of calculating inclusive fitness which are known to be unreliable. Second, I will argue that it has also lead to misinterpretation in economics of how relatedness might provide an 'exchange rate' for one person's fitness to another's. And finally, I will argue that it is behind a widespread misconception that we need inclusive fitness in order to view social behaviors as adaptations.

17.3 An unreliable heuristic

The 'simple-weighted-sum' (SWS) method of calculating inclusive fitness, famously used by Maynard Smith ([1991]), says that one can, heuristically, calculate the inclusive fitness of an organism by adding its own payoff (or fitness effects arising from some sort of social interaction) and its relative's payoff, weighted by a relatedness parameter, R (sometimes written as k).[6] This heuristic is extremely common, especially in the animal communications literature (see, e.g. Johnstone and Grafen (1992); Johnstone (1998); Nowak (2006); Taylor and Nowak (2007); Archetti (2009a,b).)

However, it is generally agreed that this is an incorrect definition (see, e.g. (Grafen (1982); Grafen (1984), Skyrms (2002); Nowak et al. (2010); Okasha and Martens (2016); Birch (2016)). For instance, this heuristic has a well-known problem with double-counting. Say we

---

[6] Compare to the definition given in section 2. Calculating inclusive fitness is often described as first stripping an organism's fitness of all the fitness effects from others, and then adding the fitness effects the organism confers on its relatives (Hamilton (1964)). By contrast, the SWS heuristic does not strip away anything and adds in all the social partner's offspring.

have two relatives, organism *A* and organism *B*, which interact and both have trait *X*. Under the SWS heuristic, when we calculate the fitness of trait *X* we count *A*'s payoff twice: once when we consider *A*'s contribution to the fitness of the trait and again (at least partially, depending on the value of *R*) when we take into account *B*'s contribution to the fitness of the trait. We similarly double-count when calculating *B*'s fitness.

Despite recognition that this calculation is incorrect, it is often viewed as a useful heuristic for estimating the inclusive fitness of traits. One intuitive argument for why this heuristic should give adequate predictions is this: if we are interested in tracking gene frequencies, adding the relatedness-weighted payoff of a relative to the focal organism's payoff means that the focal organism's genes will be passed on more often. In other words, it captures the fact that an organism in some sense cares about the payoff, or reproductive success, of its relatives and this is exactly the phenomenon that the relatedness parameter in inclusive fitness is supposed to capture.

In fact, the heuristic is often seen as preferable to explicit calculation of inclusive fitness. When payoffs are additive – i.e., when the causal effects of an organism on its social partner's fitness are the same irrespective of the type of its social partner (allowing us to just sum all these fitness effects up to determine an organism's fitness, like we did in section 2) – the heuristic correctly identifies the Nash equilibrium of a game. Further, the heuristic is easy to generalize to games where payoffs are not additive. It is difficult to use the correct calculation of inclusive fitness this type of game because it is often unclear what fitness effects an organism is causally responsible for (Okasha and Martens (2016)). In addition, it can be shown that in a game with non-additive payoffs, the heuristic at least allows one to calculate necessary, but not sufficient, conditions for something to be an equilibrium.[7] Thus, the heuristic is thought to give us an idea of the evolutionary outcomes we should expect in these more complicated models, despite the fact that it is known to have a problem of double-counting.

---

[7] In the late 1970s, the usefulness of the heuristic was debated in the context of the hawk-dove game and it was determined that heuristic sometimes gives the correct equilibrium predictions, and in other cases it lets you calculate necessary, but not sufficient, conditions for something to be an equilibrium. See Maynard Smith (1978), Grafen (1979), and Hines and Maynard Smith (1979) for this debate, and Bruner and Rubin (2020) for an overview of the conclusions.

So, the SWS heuristic is commonly used in more complex evolutionary models both because it is easier to generalize and because it captures the important feature of relatedness as generating a degree of common interest between interacting organisms. It would be fine to use this heuristic if we were to be careful in the conclusions we draw and restrict ourselves to identifying possible equilibria.

However, the heuristic is not generally used in this sort of restricted way in equilibrium analysis and is often further used in models with evolving populations to give a picture of how evolution in a population is expected to go. In dynamic analyses, the problems with the misinterpretation of relatedness become even more extreme, including both problems with inaccurate predictions and with interpretations and explanations of the evolutionary process.

To see this more clearly, let us look at the Sir Philip Sidney game, which is used in the animal communications literature to investigate biological evolution of communication between relatives. The SWS heuristic is commonly used in this context, following Maynard Smith (1991). In this game, there are two players: a sender and a receiver. The sender can be in one of two states, needy or healthy, each with some probability. Which state the sender is in is known by the sender but not the receiver. The sender has two options to try to communicate with the receiver: send a signal or not. The receiver then observes whether the signal was sent and decides whether or not to donate a resource to the sender. Donation is costly as the receiver is giving up something of value. A healthy individual will benefit from a donation, though not as much as someone in need. Individuals in a population have one strategy for when they are in the sender position (a rule for when they will signal) and one for when they are in the position of the receiver (a rule for when they will donate); we can call a combination of sender and receiver strategies a ‘total strategy’, which captures what an individual will do in each of their roles.

In this game, the receiver would not ordinarily have any incentive to donate, but the sender always wants the receiver to donate. This is one reason why it is generally thought that relatedness, or, at least partially aligning the sender’s and receiver’s interests, is important to analyzing this game. Additionally, (if it is the case that receivers generally only donate to needy individuals) senders have incentive to always try to convince the receiver they are needy. Again, we need some alignment of interests in order for ‘honest’ communication to evolve.

Consider the case where $R = 1$. In a model making use of the SWS heuristic, this translates into the organism caring as much about their social partner's payoff as they do about their own. By contrast, with relatedness appropriately conceived of a measure of correlation, $R = 1$ means that there is perfect correlation, or, that organisms always interact with someone that has the same total strategy as them. In the model with the heuristic, offspring will only signal when it is sufficiently likely their parent will be responsive to signal (that is, if there is a high enough frequency of organisms in the parent population with the strategy to only donate when they get the signal) and parents will only employ this strategy if its sufficiently likely the offspring are communicating honestly. See the red line in figure 1, which shows outcomes for a model using the SWS heuristic.[8] By contrast, when thinking about interactions with perfect correlation, this form of explanation is inappropriate; talking about likelihoods of outcomes does not make sense in this context. Because organisms are always interacting with others that has the same strategy, whatever total strategy does best against itself is the one that will evolve, as shown by the green line in figure 1.[9]

---

[8] See Bruner and Rubin (2020) for details and descriptions of the parameter values chosen. The SWS model is the same as Huttegger and Zollman (2010).

[9] Again, see Bruner and Rubin (2020) for further details, including a description of the particular model used and a defense of why it is the appropriate model to contrast with Huttegger and Zollman (2010)'s model using the SWS heuristic.

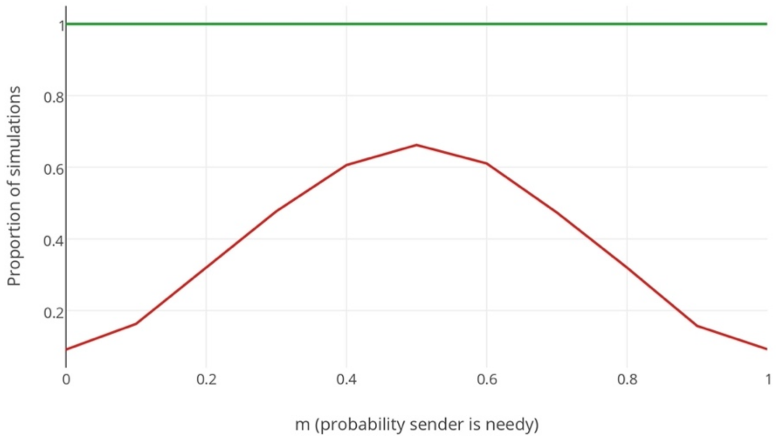


Figure 1*: Proportion of simulations which ended with populations using an honest signaling system for a model using the SWS heuristic (red) and a model which incorporates relatedness as a measure of correlation (green).*

Thus, we are given two very different pictures of how likely honest communication is for $R = 1$: if relatedness measures correlation, it is the only predicted outcome, but with the SWS heuristic, it may be very unlikely in some cases. So, the heuristic gives an inaccurate picture of which outcomes should be expected and what sorts of explanations are allowed because it ignores the fact that relatedness should measure correlation. Of course, perfect relatedness is merely the case where these issues are most apparent; a similar argument can be given for why the heuristic mis-predicts likelihoods of outcomes in cases where $R < 1$.

In sum, we have demonstrated the first problem that comes from importing ideas which are better suited for learning and decision making: thinking of relatedness as how much an organism cares about its social partner's reproductive success masks problems with and perpetuates use of an unreliable heuristic. Of course, these models using the heuristic might accurately model some scenarios in, e.g., cultural evolution, if we are considering the evolution of honest communication where the people involved care about the others they interacting with, because they have altruistic preferences or something along those lines. However, in calculating reproductive success, we ought not to replace a measure of correlation with a measure of common interest.

17.4Exchange rates

We will now turn to economics, where I will argue that thinking of relatedness as measuring 'caring' rather than correlation has led to the misuse of biological results. Until now, we have been speaking of the 'payoffs' in a game as representing fitness effects arising from some sort of social interaction. Here, we will be speaking of payoffs in terms of 'utility', in addition to their ability to represent fitness effects. In economics, payoffs generally represent what the decision maker cares about, i.e. utility they assign to outcomes, which may be different if we are thinking about economic agents versus biological entities. For instance, all a gene 'cares' about is reproducing itself, while an agent may care about another persons' well-being, fairness, etc. Of course, when we start including things like one agent caring about another, we need some idea of how much; we need some measure of how one person exchanges their own good against another person's.

Coming up with, measuring, and/or explaining these exchange rates – or understanding how *interpersonal comparisons* of utility are made – is central to much of economics, political philosophy, ethics, etc., including Ken Binmore's influential account of the evolution of our notions of justice. As Binmore explains:

> I don't suppose anyone but the most diehard of neoclassical economists denies that we actually do have standards for making interpersonal comparisons of utility, and that these are widely shared within a particular society... It is easy to guess that the origins of the capacity lie in the need for members of an extended family to recognize how closely they are related to each other (Binmore, 2005, p. 28).

That these interpersonal comparisons have their basis in recognition of family seems plausible enough on the face of it, but the argument Binmore gives rests on confusions about the role relatedness plays in the biological context.

Binmore's discussion of how kin relations gave rise to the exchange rates used in interpersonal comparisons of utility makes use of a game called the prisoners' dilemma. The prisoners' dilemma can be described as a choice between being altruistic or non-altruistic. The payoffs agents get for choosing altruism or non-altruism, depending on the trait of their social partner, are summarized in table 1a. In this game, the rational choice is always to choose *not* to be altruistic: if your social partner is an altruist you get a payoff of b rather than $b - c$ and if your social partner is not an altruist you get a payoff of 0 rather than $-c$. Before getting into Binmore's specific account, it will be useful to recount how rational decision making differs from evolution in this game, and give some terminology to help describe how economic agents could reason about correlation.

First, and most simply, when interactions in a population are random, the evolutionary prediction will be the same as the rational choice for the game: evolution will lead to a population of non-altruists, just as if the organisms were rational agents choosing their traits in order to maximize their fitness. However, interactions in a population are not always random. If there is sufficient correlation between types, the population will evolve to become composed entirely of altruists, as altruists receive the benefits from others sufficiently more often to outweigh the cost that they are paying. One might think that a rational actor should somehow take the correlation between types into consideration when deciding between traits: one should choose to be an altruist because one would be more likely to receive the benefit from interacting with another altruist. However, decision makers should not generally take correlations into account.

Table 1: Prisoners' dilemmas*. (a) a prisoners' dilemma arising from a choice of whether or not to be altruistic, and (b) the prisoners' dilemma used by Binmore (2005), which is equivalent to assuming b = 2 and c = 1, which a background fitness of 1.*

	Altruist	Not
Altruist	$b - c, b - c$	$-c, b$
Not	$b, -c$	0, 0

(a)

	Altruist	Not
Altruist	2, 2	0, 3
Not	3, 0	1, 1

(b)

To see why this is a case, let us think about the prisoners' dilemma played with a twin, as discussed by Skyrms (1994) and Sober (1998). You and your twin are apprehended by the police for committing a crime, taken into separate rooms, and each offered a deal from the police: you will get a reduced sentence if you turn state's witness and offer up evidence against your twin. You can each remain silent or turn state's witness. This relates to an altruistic action in the biological case: there is some cost to altruistically remaining silent (you could have reduced your sentence had you turned state's witness) and it also produces some benefit (withholding information means your twin can only be convicted of a lesser crime, for which the sentence is shorter).

Since you and your twin are held in separate rooms, you cannot influence each other's decisions. However, being twins, you believe that you are alike and that your choosing altruism is very good evidence your twin will as well. This means you calculate the conditional probability of their choosing to be altruist given that you yourself choose to be altruist to be high. If it is high enough, you might even choose to be altruistic because you believe the likelihood you will receive the benefit is sufficiently high so as to outweigh the cost you pay.

But the decision to be altruistic is irrational - it yields a worse payoff for no matter what your twin does. There is no reason for you to take correlations into account when evaluating the payoff consequences of your actions, as your actions will not affect anything your twin does. This sort of decision making has been referred to as 'magical thinking', as it seems to assume actions magically affect probabilities we know they cannot affect (Skyrms, 1994). This is the same sort of reasoning that occurs when talking about Newcomb's problem[10] and other related decision problems. Thus, bringing in the decision-theoretic distinction between evidential decision theory and causal decision theory will help us understand what is going on here, as it helps us understand differing intuitions in these other decision problems.

First, evidential decision theory tells decision makers to evaluate actions based on their 'news value,' or which action provides evidence that good outcomes will occur. This accounts for the intuition that one should choose to be altruistic in the above prisoners' dilemma: being an altruist is a sign that the other person will also be an altruist. However, many philosophers think that this is problematic. We should choose actions based on their consequences, yet evidential decision theory ignores the difference between cases where there is a solely

---

[10] An actor decides between only taking box B – an opaque box with either $1,000,000 or nothing depending on what a reliable (or perfectly accurate, in some variations) predictor has predicted – or also taking box A – a transparent box with $1,000 in it. If the predictor has predicted the actor will take one box (only B), there will $1,000,000 in box B. If they predicted the actor will take two boxes (both A and B), box B will be empty. Since B is already either empty or filled with money, the actor's choice cannot influence its contents.

evidential (or correlational) relationship between an act and an outcome and cases where there is a genuine causal relationship.

Causal decision theory, by contrast, only takes into account the causal consequences of an action and will choose an action based on its efficacy, or which outcomes it will produce. So, evolutionary predictions and rational decision making come apart: a rational decision maker should not generally take correlations between types into account, yet the evolutionary prediction necessarily takes correlations into account in calculations of fitness. (This point will be important again in section 5.)

However, there are many factors one might consider that could change an agent's utility function – the payoffs they assign to the different outcomes. In particular, if one cares about the success of their interactive partner, one may care not just about one's own time in jail, monetary gain, reproductive success, etc. Often, having other-regarding preferences is enough to transform the prisoner's dilemma into a game called the prisoners' delight, where the socially beneficial action is also the rational choice (e.g. table 2). Questions then arise: how do people come to have these other-regarding preferences, and how do they weight their own good versus the good of their interactive partner? According to Binmore, the ability to perform these weightings arises from our ability to exchange our own success against the success of our genetic relatives; relatedness provides an exchange rate of our utility against our social partner's.[11]

Binmore begins his argument by considering a prisoners' dilemma with a twin. He states: "When relatives play a game, the payoffs need to be identified with their inclusive fitnessess rather than their individual fitnesses" (Binmore, 2005, p. 104-5), leading him to calculate payoffs in this game by adding together both payoffs in the corresponding cell from the original prisoners' dilemma in table 1b to arrive at the payoffs in table 2. That is, Binmore is using the SWS heuristic described in section 3 to transform the payoffs of the game. Obviously, this representation of the prisoners' dilemma with a twin differs from Skyrms' and Sober's

---

[11] There are other ways to talk about relatedness as an exchange rate, e.g. Frank (1998), which are not being argued against here.

discussion of how correlation is accounted for in the prisoners' dilemma. Further, if we are thinking of relatedness as a measure of correlation, using it to transform the payoff table ought to strike one as problematic. This is because, as in standard in game and decision theory, probabilities of receiving outcomes are not included in the evaluation of outcomes. They are, instead, used to calculate the expected utility of some strategy or decision – following the causal decision theorists' reasoning, one multiplies utilities associated with possible outcomes by the probability of each outcome occurring, in a stage of analysis after the payoffs for the game have already been fixed.

Table 2: Binmore's prisoners' dilemma with a twin.

	Alt.	Not
Alt.	4, 4	3, 3
Not	3, 3	2, 2

However, it does not seem that Binmore is thinking of relatedness as a measure of correlation. He is instead identifying relatedness with caring about a social partner. In fact, he explains the transformation of payoffs as the same kind that occurs when a game is played between lovers who care about each other's outcomes (Binmore, 2005, p. 108). He then proceeds to say that, since this is a game played with a twin, we ought only to look at the payoffs along the diagonal, where both players are playing the same strategy, and model the strategic scenario as a one-person game because the twins do not choose independently, i.e. their actions are perfectly correlated (Binmore, 2005, p. 108-9). If relatedness is meant to measure correlation, why would we both use it to calculate the payoffs then to determine the likelihoods of receiving those payoffs? Even an evidential decision theorist would not argue we should take correlations into account twice.

Binmore seems to be thinking along the same lines as the quote introduced at the start of this paper, that "Relatedness might, beyond inclusive fitness, introduce additional correlation" (Zollman, 2013, p. 131), which misconstrues the nature of relatedness. This is not to say a project like Binmore's is doomed to fail. One could instead tell a story about how exchange rates emerge through interactions among family members. However, relatedness does not provide us with such an exchange rate separately from its role as a measure of correlation. That sort of thinking likely arises from describing relatedness in agential terms, as a measure of common interest.

17.5Maximizing agents

Finally, I will argue that using agential language to describe relatedness hides problems with arguments regarding the usefulness of inclusive fitness for understanding the adaptive value of social behaviors, which rest on its role in thinking of organisms as maximizing agents. The *maximizing agent analogy*, or the *heuristic of personification*, asks us to imagine what trait an organism would choose if they were an agent attempting to maximize their fitness. While there are many types of evolutionary analysis that do not require the maximizing agent analogy, e.g. predicting changes in gene frequencies, it is seen as an important for preserving Darwin's insight that selection leads to the appearance of design. As Okasha (2018) puts it, thinking about selection is not the same as thinking about adaptation and "[t]he 'fit' of an organism at which Darwin marvelled is about an organism having traits that benefit it, or further is biological goal; the agential idiom is hard to avoid here" (p. 50).[12]

So, we often think of organisms as acting 'as if' they are agents choosing traits to maximize their fitness, and therefore understand those traits as furthering biological goals. When it comes to social behavior, however, the maximzing agent analogy often cannot be straightforwardly applied. As described in the previous section, rational decision makers ought not to take correlations into account, but these correlations affect evolutionary outcomes. The traits we observe are not those that an organism 'trying' to maximize fitness would choose.[13]

---

[12] He also argues that agential thinking has advantages over talk about functions. See Okasha (2018, ch. 1) for details.

[13] Skyrms (1994); Sober (1998) both give extended arguments for this conclusion

Hamilton (1964, 1970) proposed inclusive fitness as a quantity that organisms are selected to maximize. It has since become a standard assumption that inclusive fitness is necessary in order to make sense of the appearance of design when it comes to explaining social behaviors. For instance, it is common to state that: “inclusive fitness... is a quantity that natural selection tends to cause individuals to act as if maximizing, just as Darwinian fitness tends to be maximized in the non-social case” (Grafen, 2009, p. 3137). Or, more explicitly stated, if we are going to think of organisms as maximizing agents “... doing so requires inclusive fitness” (West and Gardner, 2013, p. R579). This idea is so influential that biology students are commonly taught the principle that that natural selection leads to organisms acting as if maximizing their inclusive fitness (Grafen, 2006, p. 559).[14]

Though the usefulness of inclusive fitness does not wholly depend on allowing this sort of agential thinking, its role in the maximizing agent analogy has been seen as a major factor in explaining its popularity. As Okasha et al. (2014) put it:

> The popularity of the inclusive fitness concept in evolutionary biology arises because it allows social behaviour, even when it is individually costly, to be understood from the perspective of an individual organism ‘trying’ achieve a goal, thus preserving Darwin’s insight that selection will lead to the appearance of design in nature. (p.28)

Some claim that inclusive fitness is the only major development in our understanding of adaptations since Darwin proposed the theory of natural selection (West et al., 2011, p. 233) or that recent criticisms of inclusive fitness are irrelevant because inclusive fitness is the only concept of fitness that can play this role in explaining the appearance of design (West and Gardner, 2013, p. R582).

How would inclusive fitness allow us to use the maximizing agent analogy, when other concepts of fitness do not? West and Gardner (2013) explain:

---

[14] For instance, the following, as well as many others, all express this basic idea: Birch (2016); Gardner (2009); Grafen (2006, 2009); Okasha et al. (2014); Okasha and Martens (2016); Queller (2011); West et al. (2011); West and Gardner (2013).

> The individual does not, in general, have full control of its neighbor-modulated fitness, as parts of this are mediated by the actions of her social partners. However, the individual does have full control of inclusive fitness, as this is explicitly defined in terms of the fitness consequences for itself and others that arise out of its actions (p. R579).

Queller (2011) also notes: "This focus on what the actor can control allows us to tie into the long biological tradition of thinking of actors, or their genes, as agents. Additionally, it tells us that these agents should appear to be trying to maximize inclusive fitness" (p. 10792).

The basic argument is this: organisms are in control of their inclusive fitness because they are in control of whether they confer the benefit on their social partner, but organisms are not in control of their neighbor-modulated fitness because they are not in control of whether their social partner confers a benefit on them. That is, neighbor-modulated fitness explains the evolution of altruism in terms of 'statistical auspiciousness', or altruism happening to correlate with advantageous social neighborhoods. From a neighbor-modulated fitness point of view, if the organism could choose not to be altruistic, while keeping its social environment fixed, it would always stand to gain by doing so (Birch, 2016).

Using the terminology described in the previous section, we can say that the proponents of this *indispensability argument* are causal decision theorists: a rational actor, or maximizing agent, should only take into account the causal consequences of their decisions, not things like correlations or 'statistical auspiciousness'. Or, as West and Gardner (2013) put it: if we are going to say natural selection leads organisms to appear as if they are trying to maximize their fitness, the concept of fitness we use must be under the organism's complete control, "meaning that it is determined only by the traits and actions of the focal organism. This is because organisms can only appear designed to maximise something that they are able to control." (p. R579).

I will argue in this section that the intuitive appeal of the indispensability argument, and the reason it is so widely accepted, relies on us thinking in terms of the agential interpretation of relatedness as a measure of how much an organism cares about it social partner's reproductive success. While a goal of inclusive fitness maximization is clear when thinking of relatedness in terms of caring, it is less clear how the maximizing agent analogy is supposed to play out when thinking of relatedness as a measure of correlation.

It is important to be clear upfront what I am *not* arguing. I am not arguing that those providing this indispensability argument are misunderstanding what relatedness is, only that there is a subtle switch from talking about correlation to valuing social partners, and that this hides the role that correlation plays in the maximizing agent analogy. I am also not arguing that genealogical relatedness is unimportant to understanding the evolution of social behaviors, or that inclusive fitness is not optimized by selection, as will become clear below. Finally, I am not arguing that biologists should not interpret experimental results in terms of organisms maximizing inclusive fitness. This may in fact be the most intuitive way to understand many effects of social interactions. Rather, I am arguing that the importance of inclusive fitness as *necessary* or *the only way* to understand design in the context of social evolution is overstated and entangled with the problems described in the previous two sections.

I will make my argument in two stages below. First, I argue the idea of an organism's full or complete control is not sufficient to establish that inclusive fitness is required in order to think of organisms as maximizing agents when studying social evolution. Second, I will focus more in depth on the idea of an organism choosing a trait 'as if' they were trying to maximize some sort of fitness. I will argue that there are different ways to describe the context in which this choice is made, that not all of these decision-making contexts lead to inclusive fitness as a unique goal, and that the decision-making contexts in which inclusive fitness is a unique goal are not clearly useful for understanding the appearance of design. I will then conclude the section with some discussion of what can be said about the usefulness of the inclusive fitness framework in the maximizing agent analogy.

17.5.1 Full control

Recall that a concept of fitness is under an organism's full control when "it is determined only by the traits and actions of the focal organism." (West and Gardner, 2013). That is, the focal organism (and no other organism) must be causally responsible for the fitness consequences of the trait under consideration. For instance, the fitness consequences of an altruistic trait would include both cost and benefit terms, and both of these would need to be caused by the focal organism.

It seems intuitive to say that the production of the benefit is under the organism's control in the inclusive fitness calculation, but not in the neighbor-modulated fitness calculation, since in inclusive fitness we count the benefit the organism produces, whereas in neighbor-modulated fitness we count the benefit the organism receives. However, considering actions performed by other organisms is not generally thought to be a problem for the viewing organisms having a 'goal' of maximizing fitness.

An example from Sober (1998) will demonstrate the point. We can think about what trait an agent would want to have if they were a zebra and choosing between being a fast or slow runner. The agent would choose to be fast, maximizing fitness by escaping predators, and we can reason that natural selection will lead to a population of fast zebras. Zebras can be fast or slow runners, and which trait a zebra has determines how likely it is that it will be eaten by a predator (thus determining its fitness). In this example, the predators are considered part of the environment. They are out there eating and not eating certain zebras. As long as the trait a zebra has causally influences the likelihood that the eating or not eating will be directed towards them, we have no problem seeing their fitness as under their control.

It is the same in the case of social behavior as long as we remember that social partners are just part of the environment, out there exhibiting or not exhibiting altruistic behavior. As long as the focal organism's trait influences the likelihood that the altruistic behavior is directed toward it, we should not have any problem saying its fitness is determined by traits and actions under its control. Perhaps because the social environment includes organisms with similar traits, it is tempting to think of the organism's social partners as choosing to whether or not to bestow benefits and thus think of the benefit being under the social partner's control.[15] However, it is important to remember that, in using the maximizing agent analogy, the organism's social partners are considered part of the environment, not agents in their own right. (Our focal organism is not necessarily an agent either, but we are pretending it is in using the analogy.) The focal organism's interaction with them is just like any other interaction with the environment.

---

[15] As Lynch (2017) discusses (in the context of heritability debates), the presence of another agent who can be assigned causal responsibility can affect causal attributions (p. 36).

So, whether or not the organism actually produces the benefit ought not to factor in to our judgements about whether a benefit term is under an organism's control. For inclusive fitness, then, what matters for our judgements about the benefit term *Rb* is not whether *b* is caused by the traits and actions of the organism, but whether *R* is. But, of course, relatedness is not, in general under an organism's control. To say that it is assigns the organism casual control over something that is merely correlational. However, claims that the terms in inclusive fitness are under the organism's control focus on the costs incurred and benefits conferred, rather than on relatedness, which is generally described (in this context) as a measure of caring. Here, the agential language hides its true nature as a measure of correlation, possibly leading us to unwittingly confuse correlation with causation in talking about the *Rb* term in inclusive fitness as being something under the organism's control.

17.5.2 Decision-making context

One might object to the argument in the previous section on the grounds that I have misconstrued the decision-making context, and that relatedness is supposed to be held constant as we imagine the organism choosing a trait to maximize their fitness, since relatedness is part of the 'social context' or 'environment' of the trait. That is, one might argue, there is good reason to only focus on *c* and *b* because those result from the choices of our hypothetical maximizing agent, and there is something about the decision making context for social behaviors that makes the analogy to the non-social case inapt.

To be sure, it is not always clear what type of choice we are meant to envisage our focal organism making when using the maximizing agent analogy. This section will consider three types of choices, or three decision-making contexts, which I believe cover the possibilities that those putting forth the indispensability argument could have in mind: 1. the organism chooses a trait, as well as any associated underlying genetics, with fixed relatedness, 2. the organism chooses a trait, as well as the associated underlying genetics, without fixed relatedness, and 3. the organisms chooses a phenotype, but not the associated underlying genetics. In looking at the details of the decision-making context, we will see that each context either fails to uniquely pick out inclusive fitness as a maximization goal, or is of unclear use for understanding adaptations (which, remember, was why we were thinking of organisms as maximizing agents in the first place).

1.Choosing a phenotype and the associated underlying genetics, with fixed relatedness. In this case, we imagine our focal organism deciding its own personal phenotype, assuming that choice is accompanied both by whatever underlying genetics are associated with the trait and by the probabilities of interacting with different social partners. In other words, if the organism chooses to be an altruist, it will have a probability $P(A_j|A_i)$ of interacting with another altruist and if it chooses not to be altruistic it has probability $P(A_j|N_i)$ of interacting with an altruist. (In this section, A and N are interpreted as representing genotypes associated with altruism and non-altruism, whatever those genotypes are.) So, relatedness is held fixed as opposed to, e.g., the traits of the social partners being held fixed. That is, in this case, an organism choosing an altruistic trait makes it more likely that they interact with altruists (assuming $R = P(A_j|A_i) - P(A_j|N_i)$ is positive). In other words, we might say that the organism's decision casually influences the likelihood that they will interact with another altruist.

This is a case where it makes sense to say that the fitness consequences of the organism's choice of trait are under an organism's control. From the decision maker's point of view, their choice of trait causally influences the likelihood that they will interact with an altruist. If it is sufficiently likely that they will interact with an altruist, the rational decision is the altruistic trait, in line with the evolutionary outcome. But, both inclusive fitness and neighbor-modulated fitness work equally well for the basis of the decision maker's choice because they are both quantities that are under the organism's control. For neighbor-modulated fitness, the fact that an organism receives the benefit with a certain probability is under its control, and for inclusive fitness the fact that it is likely their social partner will share their genes is something that is under its control. This is a similar point to that made by Rosas (2010): "if controlling assortment is the clue to controlling inclusive fitness and if the organism can be credited with it, the organism controls inclusive fitness and neighbor-modulated fitness in one move" (p. 8).

Note that this reasoning holds even if the altruistic trait is conditionally expressed, as in common, for example, in explanations of worker sterility or reproductive helping. In these cases, we treat the trait in question as conditional: 'give help if such and such conditions hold', where those conditions will hold with probability p. An example of relevant conditions would be a case where an individual is stronger than another, and can take the role of a reproductive. If that condition is met, will the weaker organism stay and altruistically help the stronger reproduce?[16] A focal organism is then choosing whether it would help if the appropriate condition is met, and the consequence of choosing the altruistic trait is that there is some chance they will have to pay a cost c to confer a benefit b on their social partner.

For neighbor-modulated fitness, organisms only receive benefits when their social partner expresses the altruistic behavior (which happens when the social partner is an altruist and the conditions are met for it) and they only pay the cost when they themselves express the behavior. For inclusive fitness, an altruist only pays the cost to confer a benefit on a genetic relative when the conditions are met. So, all that happens to our initial calculations is that every term is weighted by p, the probability of the relevant conditions being met.[17] The reasoning employed by our maximizing agents does not change when the expression of the altruistic behavior is conditional. If, as is true in this decision making context, $P(A_j|A_i)$ and $P(A_j|N_i)$, and therefore R remain constant regardless of the focal organism's choice, maximizing neighbor-modulated fitness and inclusive fitness are equivalent goals.

2. Choosing a phenotype and the associated underlying genetics, without fixed relatedness. In this decision-making context, we keep fixed the traits of the organism's potential social

---

[16] This is in comparison to, for example, treating the relevant choice as between helping or reproducing. See Queller (1996) for a discussion of why it important to consider the fitness of each trait separately, i.e. choosing whether or not to accept the reproductive role when in the position to do so is one trait choice, and choosing to help or leave when not the reproductive is another. Mixing the two traits together leads to seemingly paradoxical results.

[17] More explicitly, we can calculate the neighbor-modulated fitness based on this probability:

$$\mathrm{NMF}(A_i) = P(A_j|A_i)pb - pc \qquad (1)$$

$$\mathrm{NMF}(N_i) = P(A_j|N_i)pb \qquad (2)$$

The conditional altruism trait is better when $p[(P(A_j|A_i) - P(A_j|N_i))b - c] > 0$, or $p[Rb - c] > 0$. Since the inclusive fitness of this conditional altruism is $p[Rb - c]$, the conditions for choosing altruism are the same regardless of which fitness concept we use. See Frank (1998) chapter 6 for similar calculations where the traits under consideration are tendencies of being sterile.

These calculations, of course, assume that condition (weak or strong) is independent of whether or not an organism is an altruist. If condition and trait are not independent, we could include a term that measures the influence of trait on condition (or vice versa), but this would not affect the basic conclusions we draw.

partners and let the organism decide only its own personal phenotype (while assuming that the choice of phenotype comes along with whatever underlying genetics are associated with the trait). In this case, the probability of interacting with an altruist does not change based on the focal organism's choice of trait. In other words, $P(A_j|A_i) = P(A_j|N_i)$ and so relatedness is zero. While there may be correlations in the population as a whole, the decision maker's choice does not affect *their* probability of interacting with an altruist, so (in the vein of causal decision theory) they should not take these correlations into account when making a decision – the rational choice here is not to be altruistic, whether the organism uses inclusive fitness or neighbor-modulated fitness as the measure it is trying to optimize.

One might object to this narrow focus on relatedness as correlation for a single trait; surely whole genome relatedness is the important measure to consider? That is, the choice of altruistic/non-altruistic trait is independent of the rest of the organism's genetic makeup and so if whole genome relatedness is still high, we can still think of the organism as increasing their reproductive success indirectly through increasing the reproductive success of their genetic relatives.

Whole genome relatedness is certainly relevant to a lot of reasoning surrounding social behavior, e.g. in determining whether there would be selection for a mutation at an unlinked locus which suppressed altruistic tendency. However, it is less clear how whole genome relatedness is relevant to understanding altruism as an adaptation in the context of the maximizing agent analogy. In other words, it is hard to see how we are explaining altruism as an adaptation, because in this case behavior is decoupled from inheritance – the altruistic behavior increases the reproduction of some organisms, but those organisms do not tend to have the altruistic gene more often, relative to the population average. If the point is to connect the outcomes of selection with the appearance of design, looking at whole genome relatedness when we know relatedness for the trait in question is zero breaks this connection.

3. Choosing a phenotype but not the associated underlying genetics. In this case, we might say then that the organism's choice does not determine A_i or N_i, but rather $do\text{-}A_i$ or $do\text{-}N_i$. In this case, the choice of trait is totally irrelevant to the social partners' traits or genes. This means that for neighbor-modulated fitness, the probability of receiving b for a focal organism does not depend on their choice, i.e., $P(A_j|do\text{-}A_i) = P(A_j|do\text{-}N_i)$, meaning the rational choice is $do\text{-}N_i$, to not be altruistic. Plausibly, only altering behavior would not affect any genetic relatedness between organisms, so we might think of holding fixed both $P(A_j|A_i)$ and $P(A_j|N_i)$, while also holding fixed whether the organism in question has the *A* or *N* genotype. This means that for inclusive fitness, *R* remains constant whatever the organism chooses, meaning they compare $Rb-c$ with 0, and choose to *do-*A_i.

So, it seems that, in this decision-making context, we can think of inclusive fitness as the goal of a maximize agent, where we cannot think of neighbor-modulated fitness in the same way.[18] But does this get us the conclusion that inclusive fitness allows us to see how evolution leads to the appearance of design? The problem is that imagining an organism in this decision-making context is not of obvious use for understanding the outcomes of natural selection, which describes changes in gene frequencies or traits over time.

Similar to the issues with using whole genome relatedness when we know relatedness for the trait of interest is zero, we again have a case where behavior is decoupled from inheritance – the altruistic behavior may increase some organisms' reproductive success, but those organisms do not tend to pass on altruistic genes. Here is one consequence that demonstrates the oddity of the maximizing agent analogy in this case: we would conclude that it is evolutionarily advantageous for organisms without altruistic genes to be altruists. These organisms can maximize their inclusive fitness by providing relatives with the benefit *b*, it just so happens that they are helping their relatives pass on non-altruistic genes. While we might

---

[18] For a more explicit argument, we can calculate the neighbor-modulated and inclusive fitness for organisms who choose *do-*A_i or *do-*N_i. Since these choices are independent of genetics, we calculate the probability of receiving a benefit depending on whether the focal has the underlying *A* or *N* genotype, with probability $P(A_i)$ or $P(N_i)$:

$$\text{NMF}(do\text{-}A_i) = [P(A_j|A_i)P(A_i) + P(A_j|N_i)P(N_i)]b - c \tag{3}$$

$$\text{NMF}(do\text{-}N_i) = [P(A_j|A_i)P(A_i) + P(A_j|N_i)P(N_i)]b \tag{4}$$

Comparing these two, we can see that the neighbor modulated fitness of *do-*A_i is always *c* less than *do-*N_i, and so the rational choice is not to be altruistic. By contrast, the inclusive fitness of the altruistic action is $Rb - c$, where:

$$R = [P(A_j|A_i) - P(A_j|N_i)]P(A_i) + [P(N_j|N_i) - P(N_j|A_i)]P(N_i) \tag{5}$$

and of the non-altruistic action is 0. Assuming relatedness is high enough, the rational choice is to be altruistic. (Note that often $P(A_j|A_i) - P(A_j|N_i) = P(N_j|N_i) - P(N_j|A_i) = R$.)

conclude that, in some sense, altruism is advantageous, it is unclear how this helps us say anything about the *evolution* or *adaptive value* of altruism.[19]

17.5.3 Maximizing fitness

It is worth repeating that I am not claiming proponents of the indispensability argument misunderstand what relatedness is. Rather, I am suggesting that in providing the intuitive gloss of relatedness as a measure of caring, they push relatedness to the side in considering the consequences of an organism's traits in order to focus on the costs and benefits. In doing to, correlation is allowed to sneak in, disguised as part of (or a weighting of) the benefits to social partners, which are considered to be wholly under the organism's control. Describing relatedness as how much an organism cares about its social partner's reproductive success masks difficulties with using inclusive fitness in the maximizing agent analogy and makes the indispensability argument appear much more straightforward than it actually is: the organism chooses whether to pay a cost c to a confer a benefit b on a social partner depending on how much it values that social partner's reproductive success, R.

However, when we try to give the same argument while describing relatedness as a measure of correlation, things become more complicated. The previous section gave a characterization of when we can expect inclusive fitness to be an organism's unique maximization goal, in terms of how we conceptualize the decision-making context of our focal maximizing agent organism. I then argued that those cases in which inclusive fitness is a unique goal are cases where a choice of trait is decoupled from inheritance, which seems to make the maximizing agent analogy ineffective for reasoning about adaptations or evolution leading to the appearance of design. Proponents of the indispensability argument may try to defend the usefulness of the maximizing agent analogy in this kind of decision-making context, but it seems at odds with

---

[19] This is somewhat similar to a point made by Okasha (2018, ch. 1), who argues that agential thinking surrounding decisions arising from flexible behavior is aimed at proximate causes (or, providing explanations in terms of physiology or the environment), whereas the maximizing agent analogy aims at ultimate causes (providing evolutionary explanations).

how the analogy is usually conceptualized (in terms of attempting to explain ultimate, rather than proximate causes of traits) and would leave us with a more complicated story to tell than the simple, intuitive maximization of $Rb-c$ we get when relatedness is thought of as a measure of how much the organism cares about their social partner's reproductive success.

17.6Conclusion

It is tempting to think of relatedness in agential terms, as a measure of 'common interest' or 'caring'. If an organism's goal is to pass on its genes, and its social partner is likely to share those genes, then in some sense the organism 'wants' the social partner to succeed or 'cares' about its reproductive success. Relatedness measures how likely it is that organisms share genes, relative to the rest of the population, so it does, in a sense, give us idea of the degree of common interest between them. Noticing this can be useful as a way to intuitively understand how correlation affects the evolution of social behaviors.

However, we can take this reasoning too far and supplant or replace our understanding of relatedness as a measure of correlation with an understanding of it in these agential terms. I have argued that a lack of caution in reasoning about relatedness has led to reliance on unreliable heuristics (section 3) and misunderstandings (section 5) in biology, as well as to misuse of biological concepts in economics (section 4). Instead, we ought to acknowledge the limitations in borrowing ideas from economics like 'degree of common interest' and importing them into the biological context. These ideas are appropriate to incorporate into evolutionary dynamics when talking about learning among (boundedly) rational agents, but incorporating them into biological evolution can lead us astray.

Acknowledgments

I have presented and received helpful feedback on research relevant to arguments in this paper from audiences at: POBAM 2016, the 2016 PSA, the Social Dynamics seminar at UC Irvine, University of Notre Dame, The Generalized Theory of Evolution (Dusseldorf), British Society for the Philosophy of Science 2018, The Philosophy of Social Evolution (LSE), Carnegie Mellon University, and Evolutionary Thinking Across Disciplines (Paris). Thanks also to Justin Bruner, Jonathan Birch, Brian Skyrms, and Simon Huttegger for helpful discussions related to various parts of this paper.

References

Archetti, M. (2009a). Cooperation and the volunteer's dilemma and the strategy of conflict in public goods games. *Journal of Evolutionary Biology 22*, 2129– 2200.

Archetti, M. (2009b). The volunteer's dilemma and the optimal size of a social group. *Journal of Theoretical Biology 261*, 475–480.

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.

Binmore, K. (2005). *Natural justice*. Oxford university press.

Birch, J. (2016). Hamilton's two conceptions of social fitness. *Philosophy of Science 83*(5).

Bruner, J. P. and H. Rubin (2020). Inclusive fitness and the problem of honest communication. *The British Journal for the Philosophy of Science 71*(1), 115–137.

Dugatkin, L. A. (2007). Inclusive fitness theory from darwin to hamilton. *Genetics 176*(3), 1375–1380.

Frank, S. A. (1998). *Foundations of social evolution*. Princeton University Press.

Gardner, A. (2009). Adaptation as organism design. *Biology Letters 5*(6), 861–864.

Grafen, A. (1979). The hawk-dove game played between relatives. *Animal behaviour 27*, 905–907.

Grafen, A. (1982). How not to measure inclusive fitness. *Nature 298*(29), 425– 426.

Grafen, A. (1984). Natural selection, kin selection and group selection. In J. Krebs and N. Davies (Eds.), *Behavioural Ecology* (2 ed.). Oxford: Blackwell Scientific Publications.

Grafen, A. (2006). Optimization of inclusive fitness. *Journal of Theoretical Biology 238*(3), 541–563.

Grafen, A. (2009). Formalizing darwinism and inclusive fitness theory. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 364*(1533), 3135–3141.

Hamilton, W. D. (1963). The evolution of altruistic behavior. *The American Naturalist 97*(896), 354–356.

Hamilton, W. D. (1964). The genetical evolution of social behavior i and ii. *Journal of Theoretical Biology 7*, 1–16.

Hamilton, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model.

Hines, W. G. S. and J. Maynard Smith (1979). Games between relatives. *Journal of Theoretical Biology 79*(1), 19–30.

Huttegger, S. and K. Zollman (2010). Dynamic stability and basins of attraction in the sir philip sidney game. *Proc. R. Soc. B 277*, 1915–1922.

Johnstone, R. (1998). Efficacy and honesty in communication between relatives. *The American Naturalist 152*, 45–58.

Johnstone, R. and A. Grafen (1992). The continuous sir philip sidney game: a simple model of biological signaling. *Journal of Theoretical Biology 156*, 215–234.

Lynch, K. E. (2017). Heritability and causal reasoning. *Biology & Philosophy 32*(1), 25–49.

Marshall, J. A. (2015). *Social Evolution and Inclusive Fitness Theory*. Princeton University Press.

Maynard Smith, J. (1978). Optimization theory in evolution. *Annual Review of Ecology and Systematics 9*, 31–56.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge university press.

Maynard Smith, J. (1991). Honest signaling, the philip sidney game. *Animal Behavior 42*, 1034–1035.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *science 314*(5805), 1560–1563.

Nowak, M. A., C. E. Tarnita, and E. O. Wilson (2010). The evolution of eusociality. *Nature 466*(26), 1057–1062.

Okasha, S. (2018). *Agents and Goals in Evolution*. Oxford University Press.

Okasha, S. and J. Martens (2016). Hamilton's rule, inclusive fitness maximization, and the goal of individual behaviour in symmetric two-player games. *Journal of evolutionary biology*.

Okasha, S., J. A. Weymark, and W. Bossert (2014). Inclusive fitness maximization: An axiomatic approach. *Journal of theoretical biology 350*, 24–31.

Queller, D. C. (1996). The measurement and meaning of inclusive fitness. *Animal Behaviour 51*(1), 229–232.

Queller, D. C. (2011). Expanded social fitness and hamilton's rule for kin, kith, and kind. *Proceedings of the National Academy of Sciences 108*(Supplement 2), 10792–10799.

Rosas, A. (2010). Beyond inclusive fitness? on a simple and general explanation for the evolution of altruism. *Philosophy & Theory in Biology 2*.

Rubin, H. (2018). The debate over inclusive fitness as a debate over methodologies. *Philosophy of Science 85*(1), 1–30.

Skyrms, B. (1994). Darwin meets the logic of decision: Correlation in evolutionary game theory. *Philosophy of Science 61*(4), 503–528.

Skyrms, B. (2002). Altruism, inclusive fitness, and the logic of decision. *Philosophy of Science 69*, S104–111.

Sober, E. (1998). Three differences between deliberation. *Modeling rationality, morality, and evolution* (7), 408–422.

Taylor, C. and M. A. Nowak (2007). Transforming the dilemma. *Evolution 61*(10), 2281–2292.

West, S. A., C. El Mouden, and A. Gardner (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior 32*(4), 231–262.

West, S. A. and A. Gardner (2013). Adaptation and inclusive fitness. *Current Biology 23*(13), R577–R584.

Zollman, K. (2013). Finding alternatives to the handicap principle. *Biological Theory 8*, 127–132.

Part IV: How can evolutionary approaches or the target field be amended?

Chapter 18

From the Modern Synthesis to the Inclusive Evolutionary Synthesis: an Einsteinian revolution in evolution

Étienne Danchin[i]

Abstract

After reviewing the evidence that led many evolutionary biologists to highlight the necessity to have the Modern Synthesis of Evolution evolve towards a new more integrative framework, I discuss characteristics that should prevail in this endeavour. I sketch a pathway towards the new synthesis by building the new synthesis around inheritance mechanisms (i.e. parent-offspring resemblance) as this is a keystone concept of evolutionary approaches since Darwin and Wallace. For this, we must incorporate all known mechanisms of inheritance into the new synthesis. One way to facilitate this effort would be to enshrine this inclusive ambition in the name of the new synthesis, for instance by calling it the "Inclusive Evolutionary Synthesis (IES)". Historically, the Modern Synthesis unified two broad fields of what Mayr called "evolutionary biology" (as opposed to "functional biology"). Building the IES will imply merging both functional and evolutionary biology into a single integrative framework. To my opinion, this constitutes "The" major challenge in this endeavour as these two fields have been independent for decades. I finally compare the current change in the general evolutionary framework to the one that happened when astrophysics moved from Newton to the special and then general relativity a hundred years ago.

---

[i] Laboratoire Évolution and Diversité Biologique (EDB UMR 5174), CNRS, Université de Toulouse, IRD, 118 route de Narbonne, F-31062 Toulouse cedex 9, France. Email: etienne.danchin@univ-tlse3.fr

18.1Introduction

At the turn of the third millennium, the multibillion-dollar project of completely sequencing the human genome became a major landmark in evolutionary sciences. That project fostered major advances in DNA sequencing technologies leading to the emergence of high-throughput sequencing and many new molecular methods enabling the very fine description of genetic variation. Admittedly, this opened a quasi-unexplored vast domain of research, and evolutionary biologists were excited about having access to the entire human genetic sequence. Everything was there, we thought, written in that fantastic sequence that would allow us to understand and alleviate our diseases! There was reason to rejoice. Although this first sequencing had cost around 3 billion dollars, it was worth it, because we were finally on the verge of associating actual genetic variation with phenotypic variation! This had been the hidden dream of every geneticist and evolutionary biologist for a long time. After all, the uncovering of the complete human genetic sequence opened up infinite possibilities for associating genetic variation (i.e. variation in DNA sequence) with trait variation, thus setting the stage for developing new therapies to correct so-called genetic diseases. At last, such dreams might become reality!

Researchers soon developed such long-awaited methods, which involved DNA sequencing to search for variants, either in whole genomes, sub-genomes or in millions of fragments scattered throughout the genome. The detected variation could then be correlated with variation in phenotypic traits. Although we know that co-occurrence does not demonstrate causality, documenting associations between genetic variation and phenotypic variation would provide an indeed non-compelling, but nonetheless significant argument for a link between DNA sequence variation and phenotypic variation, a link that is at the heart of genetics. For example, by comparing sequences across many individuals, it became possible to characterize variants at the nucleotide level and identify 'Single Nucleotide Polymorphism', or SNPs, i.e. specific positions in the DNA sequence that show some variation. These SNPs are a formidable tool because thousands or even millions of them can be detected, for example, in a study of human embryonic stem cells and foetal fibroblasts, which involve more than 12 million SNPs spread throughout the genome (Exactly 12 236 967 SNP, Xia *et al.* 2012). This type of very promising approach was dubbed 'Genome Wide Association Studies' or GWAS.

All these fantastic developments were intended to improving our understanding of heredity, a concept that sanctions the ubiquitous fact that offspring resemble their parents more than any non-relative conspecific. The study of the mechanisms responsible for that resemblance (i.e. inheritance) has been, and still is, at the centre of biology. A classic definition is that heredity depicts patterns of parent-offspring resemblance.

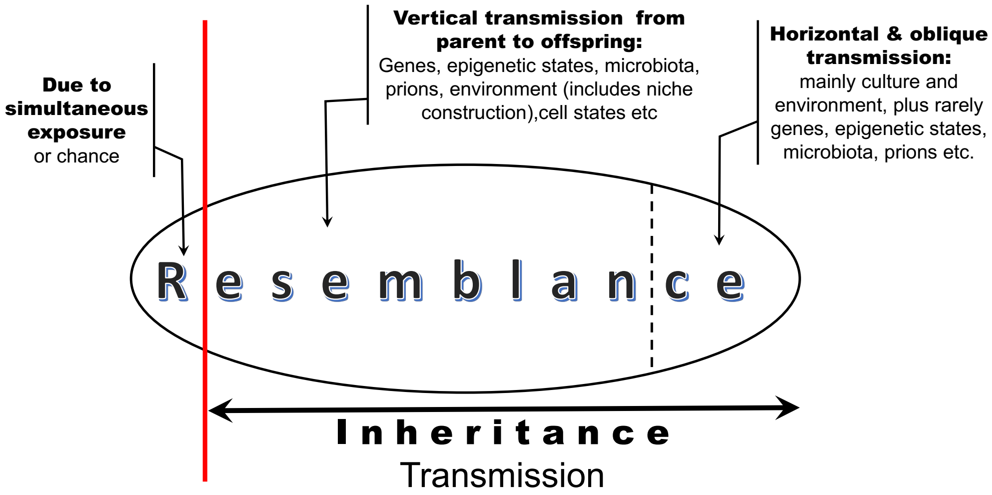


Figure 1: Transmission and phenotypic resemblance. *The ellipse represents all cases of within population resemblance. Left of the red line are cases that emerge by chance or that result from simultaneous exposure by which environmental stresses affect the soma affect the soma and the germline (or developing embryos) of individuals. Such cases do not involve transmission. The rest of resemblance may result from vertical transmission (from parents to offspring) or from horizontal or oblique transmission among individuals of a population and involve the transmission of some information among population members.*

This classic definition, however, focuses on resemblance resulting from vertical transmission, i.e. from parent to offspring and hence among relatives. But, non-vertical transmission also exists, in genes (though very rarely in eukaryotes), and much more commonly in cultural inheritance where horizontal or oblique transmission occur, hence fostering resemblance among non-relative members of the same population (i.e. a group of interacting individuals). A more general definition of heredity could thus be "*patterns of resemblance that result from the transmission of some information among individuals*", which can be shortened int "*transmitted resemblance*" as I do below. Information here is understood in a broad meaning. It includes information with a well-defined avatar, (*i*) as in genetic (DNA sequence), (*ii*) epigenetic information (epigenetic marks), or (*iii*) in the form of a stable molecule shape and function (prions and chaperon). It can also include information with avatars that are less easy to define, such as (*iv*) the transmission of cellular states beyond epigenetic states, or (*v*) that of the environmental state, or (*vi*) the transmission of microbiota. Finally, it also includes (*vii*) information transmitted among individuals that has no real avatar as through social learning and cultural inheritance. *The important point is that such transmission should lead to resemblance that is stable intergenerationally (Figure 1).*

In this paper I use the above general definition of heredity that focusses on patterns of resemblance among relatives or non-relatives, and the term "inheritance" to depict the mechanisms that produce those patterns of resemblance.

The mainstream view is that inheritance mainly, if not exclusively, results from the transmission of the information engraved into the sole DNA sequence. That assumption is at the heart of the view of inheritance that characterises the Modern Synthesis of Evolution, which has been the most widely accepted framework to study evolution since its establishment by Fisher (1930), Wright (1930, 1931), Haldane (1932 [1990]), Hamilton (1964), Williams (1966 [2019]), (Dobzhansky 1973), (Mayr 1982), (Simpson 1984) and Dawkins (1976, 1982) among many others, in the decades around the middle of the 20th century; for a comprehensive and scholarly review see Ågren (2021).

Implicitly, all the new approaches described above using the human genome data that became possible owing to the advent of high-throughput DNA sequencing were intended to establish the success of the Modern Synthesis once for good. But, it soon appeared that it would be far from being the case, far from it. In particular, many human diseases known as "genetic" for the sole reason that they are transmitted (i.e. offspring of ill parents are likely to develop the same diseases) cannot be explained by purely sequencic information (i.e. information engraved into the DNA sequence). This is the case, for instance, of obesity, diabetes, dependence to drugs, autism, depression, ill parental behaviour etc. As a matter of facts, evidence accruing from all domains of biology show that the gene-centric vision of inheritance is not sufficient to explain the full complexity of life and reveals the existence of a suite of non-genetic mechanisms of inheritance, of incredible variety and sophistication. These involve the transmission of epigenetic, cell, ecologic, cultural, prions, chaperon molecules states, as well as the transmission of the microbiota with which each multicellular organism is living.

Here, I will adopt a historical perspective to review briefly some of the evidence that keep on accruing and reveal that the gene-centric vision of inheritance cannot explain the full complexity of living organisms. Ultimately, following many others (Pigliucci 2005, Champagne 2008, Bonduriansky and Day 2009, Champagne and Curley 2009, Danchin and Wagner 2010, Franklin and Mansuy 2010, Skinner *et al.* 2010, Danchin *et al.* 2011, Skinner 2011b, a, Bohacek and Mansuy 2015, Muller 2017, Wang *et al.* 2017, Bonduriansky and Day 2018, Danchin *et al.* 2018, Danchin *et al.* 2019b, Danchin 2022), I advocate the necessity for a new integrative synthesis to incorporate all components of inheritance, whether genetic or not, into a more integrative evolutionary synthesis. To flag that ambition, I recommend using the term “inclusive” to qualify that new synthesis, in what I call the Inclusive Evolutionary Synthesis (or IES). I will explain in what way this IES vision of evolution has a better potential to apply Darwinism to other domains of biology such as animal culture and language evolution. I also discuss some of the main challenges to overcome to make the new synthesis become effective, and widely endorsed.

18.2 Accruing evidence calling for a new synthesis

18.2.1. The case of the missing heritability

I will start with the case of the missing heritability, because it is highly illustrative of our collective focus on the role of the information encoded in the DNA sequence in heredity. It is that view that led me and colleagues to coin the term ‘*sequencic*’ to qualify the fact that we are all implicitly convinced that DNA sequence is the *only* heritable information that produces parent-offspring resemblance (Danchin *et al.* 2019a). We were taught that idea repeatedly during all our training, as well as in the media. By extension, in this paper, I will use ‘genetic’ and ‘sequencic’ as equivalent terms, as it now matches the mainstream usage of the term genetic in all components of the society, and particularly among biologists of all sorts. Consequently, by non-genetic (or non-sequencic) inheritance I mean any mechanism of resemblance among relatives or interacting non-relatives that does not rest on variation in DNA sequence (Danchin and Wagner 2010, Danchin *et al.* 2011).

After the advent of high-throughput sequencing, it soon appeared that GWAS approaches led to unexpectedly low estimates of heritability compared to populational estimates obtained with the classical method measuring parent-offspring resemblance at the population scale, for instance in ecology or in epidemiology. In 2008 already, Brendan Maher (2008) reviewed the disturbing evidence for that enigma that begged for an explanation because heritability is one of the keystones of the conceptual edifice that constitutes the Modern Synthesis as it quantifies parent-offspring resemblance without which no evolution can occur. Despite big efforts to control for a series of potential environmental effects, population estimates of heritability appeared invariably much higher than the newly obtained estimates in GWAS. This was disturbing because GWAS estimates directly results from particularly fine grain descriptions of DNA sequence variation often on big samples. Maher dubbed that enigma, the "case of the missing heritability", which marked the beginning of an abundant literature seeking to explain the enigma. In a bibliographic search in late September 2021, I found that the expression "missing heritability" was one of the themes of more than 1,990 articles according to the Web of Knowledge (leading to a peak of more than 10,000 citations in 2020), and more than 21,400 hits according to Google Scholar. Obviously, the case of the missing heritability has become a hot topic.

The existence of missing heritability is illustrated usually by the ratio of the heritability estimate by GWAS to that obtained for the same trait by classical methods at a population scale. Both methods measure correlations. However, the former (GWAS) is unquestionably based on the occurrence of sequencic variation. It quantifies the statistical association between sequencic variation and trait expression. Contrastingly, the classical measurements of heritabilities just quantify the statistical level of parent-offspring resemblance, *in the absence of any sequencic information*. In other words, it does not incorporate any molecular information.

Interestingly, most articles of the ever-growing massive piece of literature on missing heritability focus on the limitations of estimates of heritabilities via GWAS. These limitations are undisputable, and my goal here is not to review them. However, the fact that the ratio of GWAS to classical estimate of heritability is low can have two non-exclusive causes. It might be that the numerator (GWAS estimate) is underestimated, or that the denominator (classic estimates) is overestimated, or both. Therefore, instead of focusing *only* on correcting the potential underestimation of GWAS approaches, we should *also* explore the possibility that estimates of heritabilities by the classical methods are in fact largely overestimated because it is quasi impossible to control for all potential mechanisms of non-genetic parent-offspring resemblance (Danchin 2013, Bourrat *et al.* 2017, Bourrat and Lu 2017).

For instance, Manolio *et al.* (2009), the most cited paper on missing heritability (with over 8,200 citations in 12 years), quasi ignores the potential role of non-genetic inheritance. With one small exception, it merely focuses on the reasons why GWAS heritabilities are underestimated. Its Box 3 illustrates my point in that it proposes a research agenda for the future encompassing nine general approaches to solve the missing heritability problem. That box only alludes to the role of epigenetic inheritance in passing when discussing the eighth approach dealing with the study of gene-environment interactions. That eighth approach contains seven items, one of which being the study of the potential role of epigenetics in inheritance. In other words, on top of giving to epigenetic a marginal role, that highly cited research agenda almost ignores the potential role of most of the non-genetic processes of inheritance that have been identified during the last decades, namely, cultural, environmental, cytoplasmic and cell inheritance, as well as the inheritance of the microbiota, of chaperon molecules and prions, and niche construction (reviews in Odling-Smee *et al.* 2003, Danchin and Wagner 2008, Danchin and Wagner 2010, Danchin *et al.* 2011, Danchin 2013, Lu and Bourrat 2017, Danchin *et al.* 2019b). Unfortunately, this attitude is highly revealing of the general denial of the importance of non-genetic inheritance altogether in evolution.

Since 2009, two important discoveries have made even more salient the necessity to contemplate the possibilities that classic estimates of heritabilities are overestimated.

First, improvements in the statistical methodology to estimate heritability with GWAS data have led to heritability measurements closer to those obtained by populational methods. However, even after accounting for several processes correcting for the underestimation of heritabilities in GWAS, the mismatch between GWAS and classical approaches still remains much bigger than anticipated, and some studies strongly suggest that, at least for some traits, the mismatch will *never* be reduced to zero (e.g. López-Cortegano and Caballero 2019).

Second, evidence for a massive and ubiquitous role of non-genetic inheritance has accumulated at a fast pace. Today, the list of studies documenting such processes is far too long to cite all existing evidence. This is what led me and Richard H. Wagner to propose the concept of 'inclusive heritability' to depict the degree of transgenerational resemblance among relatives or non-relatives *whatever the mechanism responsible for it* (Danchin and Wagner 2010, Danchin *et al.* 2011). A partial list of striking case studies is (Francis *et al.* 1999, Anway *et al.* 2005, Skinner and Anway 2005, Crews *et al.* 2007, Skinner *et al.* 2010, Guerrero-Bosagna and Skinner 2012, Dias and Ressler 2014), and major reviews are (Pigliucci 2005, Champagne

2008, Bonduriansky and Day 2009, Champagne and Curley 2009, Danchin and Wagner 2010, Franklin and Mansuy 2010, Skinner *et al.* 2010, Danchin *et al.* 2011, Skinner 2011b, a, Bohacek and Mansuy 2015, Muller 2017, Wang *et al.* 2017, Bonduriansky and Day 2018, Danchin *et al.* 2018, Danchin *et al.* 2019b, Danchin 2022). Altogether, the weight of evidence strongly supports the idea that non-genetic inheritance is all over the place, suggesting that most if not all populational estimates of heritability are capturing the effect of some non-genetic components of parent-offspring resemblance. In other words, such overestimated heritabilities are in fact likely to correspond to inclusive heritability (Danchin and Wagner 2010, Danchin *et al.* 2011) rather than to purely genetic (i.e. sequencic) heritability as usually understood.

I took the time to develop the case of the missing heritability because, by showing the extent to which our purely sequencic view of inheritance is not sufficient to explain the complexity of life, it stresses the urgency for evolutionary biology to establish a new highly integrative synthesis incorporating all dimensions of inheritance, whatever the mechanisms responsible for resemblance. It is the ultimate goal of this chapter to try to convince tenants of the new synthesis of the necessity to integrate *all dimensions* of non-genetic (i.e. non-sequencic) inheritance into the new synthesis.

18.2.2. Integrating cultural inheritance

Historically, decades before the emergence of the missing heritability enigma, human sciences first brought indirect evidence that other forms of inheritance exist, and that reducing inheritance to the sole transmission of the DNA sequence was too much of a reductionist view. Initially this field only concerned humans, but as always, it soon percolated through a vast array of taxa. Starting in the late 1970s, arguments were published showing that there is a potential conflict between cultural and genetic inheritance that could bring populations to equilibria that would be impossible to reach with the sole inheritance of sequencic information (Cavalli-Sforza and Feldman 1981, Boyd and Richerson 1983, Cavalli-Sforza and Feldman 1983, Feldman and Cavalli-Sforza 1984, Boyd and Richerson 1985, Feldman and Cavalli-Sforza 1989, Boyd and Richerson 1995, Richerson and Boyd 2005).

Since then the taxonomic scope of cultural inheritance has been extended far beyond the sole humans, over most vertebrates (Fish, birds and mammals, including various ungulates, cetaceans and primate: Whiten *et al.* 1999, Whiten 2005, Allen *et al.* 2013, Krützen *et al.* 2014, Aplin *et al.* 2015, Jesmer *et al.* 2018), and at least some invertebrates (insects, Alem *et al.* 2016, for the within hive transmission of a new foraging technique in bumblebees, or Danchin *et al.* 2018, for the transmission of sexual preferences in the fruit fly. Recent review in Whiten 2021).

More generally, the study of social learning, which constitutes the basic process for the emergence of cultural traditions (the most striking marker of an ongoing cultural process), has flourished in all vertebrates classes and beyond at least in insects. Social learning concerns a vast range of fitness enhancing processes including foraging, predator avoidance, laying habitat selection, and mate choice (Galef 1990, Heyes 1994, Heyes and Galef 1996, Danchin *et al.* 1998, Danchin and Wagner 1999, Doligez *et al.* 1999, Wagner *et al.* 2000, Doligez *et al.* 2002, Brown and Laland 2003, Doligez *et al.* 2003, Danchin *et al.* 2004, Griffin 2004, Coolen *et al.* 2005, Danchin *et al.* 2005, Leadbeater *et al.* 2006, Leadbeater and Chittka 2007, Chapman *et al.* 2008, Danchin *et al.* 2008a, Danchin and Wagner 2008, Sarin and Dukas 2009, Rendell *et al.* 2010, Slagsvold and Wiebe 2011, van de Waal *et al.* 2013, Varela *et al.* 2018, Whiten 2018, Danchin *et al.* 2022). Altogether, the ubiquity of social learning shows that the main prerequisite of cultural inheritance exists in a vast array of organisms, suggesting that the cultural phenomenon is potentially a widespread process of inheritance at least in brained organisms. Furthermore, theoretical considerations and empirical studies showed that the cultural process can participate to resemblance among relatives and non-relatives and significantly affect the genetic structuring of natural populations (Laland 1994, Laland *et al.* 2010). Hence, the existence of cultural inheritance acts through two separate processes, one that brings another form of non-genetic inheritance, and one that also has the potential to significantly affect genetic inheritance. The latter process, mainly results from the fact that cultural transmission can also be oblique and horizontal.

Concerning the potential role of the cultural component of non-genetic inheritance, with colleagues we have been using the logic of quantitative genetics since 2004 to study the potential role of the cultural component in inheritance (Danchin *et al.* 2004). I generalised that idea in subsequent papers (Danchin and Wagner 2010, Danchin *et al.* 2011, Danchin 2013), some of which proposing new methods to separate the genetic from the non-genetic components of inheritance (Danchin *et al.* 2013). This approach led us to propose a new and testable definition of culture that contrasts with previous definitions that mainly focused on the pattern of persistent behavioural variation existing among populations of the same species. Our new definition constitutes a kind of toolbox to identify cultural processes in any kind of organism.

The definition has two components. The first part is rooted in quantitative genetics. It concerns the existence of patterns of behavioural variation among populations (usually called traditions): "culture is the part of phenotypic variation that is inherited by social learning (i.e. learning from others)" (Danchin and Wagner 2010, Danchin *et al.* 2011). The second part identifies 4 criteria (Danchin and Wagner 2008, Danchin and Wagner 2010, Danchin *et al.* 2011), that later became 5 criteria (Danchin *et al.* 2018) concerning the properties of social learning and that need to be met simultaneously to ascertain that a trait is transmitted culturally. Criteria are, the trait 1) should be learned socially, 2) be transmitted often from older to younger individuals, 3) by memorised for sufficient time for having a chance to be copied, and 4) be trait- rather than individual-based. Finally, 5) it should incorporate a repair mechanism such as conformity, punishment, or digitalisation (Danchin *et al.* 2018). Together, these five criteria can generate any type of transmission, be it vertical, horizontal or oblique depending on the kinship between learners and the model from which they learn. We tested that definition in the fruit fly (*Drosophila melanogaster*), a non-social insect that constitutes a somewhat silly species to test for the existence of animal culture (Danchin *et al.* 2018). Our study dealt with the transmission of sexual preferences and involved a series of experiments coupled with some modelling to test for the potential of the observed properties of social learning to generate local traditions in mating preferences. We concluded that that species has all the cognitive capacities to transmit its sexual preferences socially in a way that can potentially foster the emergence of local collective preferences (i.e. local traditions Danchin *et al.* 2018).

18.2.3. Integrating epigenetics

The next domain of biology to climb on the train for the new synthesis was epigenetics, a domain that claimed that, by tending to reduce inheritance to the sole transmission of the DNA sequence, the Modern Synthesis of Evolution was missing major processes of evolution. That claim actually integrated and reformulated a series of previous claims that had been rampant ever since at least the discovery of the genetic code in the 1950s, which marked the full establishment of the Modern Synthesis in its current understanding.

A pioneer in this domain is Eva Jablonka who started to call for a revision of the Modern Synthesis in the 1990s (for instance among many others: Jablonka and Lamb 1989, 1995, Jablonka *et al.* 1995, Lachmann and Jablonka 1996, Jablonka *et al.* 1998, Avital and Jablonka

2000, Jablonka and Lamb 2010, Jablonka 2013). She claimed that the Modern Synthesis did not satisfactorily explain the complexity of life. Her book “Evolution in four dimensions” with Marion J. Lamb identified four major sources of inheritance, genetic, epigenetic, behavioural and symbolic (Jablonka and Lamb 2005, for a good review on that book see Pigliucci 2005). I am always reluctant to use processes meant to capture the specificities of our species because the history of science as recurrently shown that what we first think as unique to humans in fact also exists in many other species, often in a more basic but nonetheless efficient form. Here, I do not see the reasons why Jablonka and Lamb (2005) isolate what they call symbolic from what they call behaviour (which in fact corresponds to cultural inheritance). I would indeed argue that the waggle dance of the bees is symbolic in essence, and I would not be surprised that we discover other examples of symbolic communication in the future. Separating symbolic from cultural transmission can only weaken their major argument that there are *several* inheritance systems. I personally distinguish more than four inheritance systems (see Figure 6 of Danchin *et al.* 2019b), but this is secondary here.

Today, although the domain of the role of epigenetics in inclusive heritability progressed significantly in the last decades, owing to many studies in humans and its rodents counterparts among other biological models the conclusion of Bossdorf *et al.* (2008) that we are still lacking studies on the non-genetic inheritance of ecologically relevant traits in nature remains largely valid. This is not necessarily a problem at this stage, as the technicity required by studies of epigenetics remains relatively incompatible with field studies. However, this is one of the necessary steps for the full future integration of epigenetic inheritance into the evolutionary synthesis.

18.2.4. Integrating development

In parallel, the evo-devo approach, which emerged in the 1980s, insisted on the importance of development as a major actor of evolution (Arthur 2004, Muller 2007). In the early 2000s, Mary-Jane West-Eberhard (2003) was one of the pioneers highlighting the necessity of revising the Modern Synthesis. In her magisterial book she reviewed many processes of development that may significantly affect evolution by generating phenotypic variance, particularly for the part of these phenomena that leads to inclusively heritable variation. Although at that time she barely used the expression 'non-genetic inheritance', she was in fact documenting and discussing such kind of processes.

One year later, Wallace Arthur (2004) highlighted the importance of early development as *thee* major moment in the lifecycle of all organisms during which most phenotypic variation is produced. His claim was that development is a major component of the evolutionary process. He patiently and didactically exposed a series of reasons why we should consider development as the second pillar of evolutionary processes alongside natural selection. In particular, he insisted on the fact that organisms are not born adult. Organisms develop from a single cell into adults, and that development produces the vast majority of phenotypic variation. He thus called for the incorporation into the new synthesis of the evolution of development as *thee* generator of the phenotypic variation on which selection or drift can act. More recently, (following predecessors such as Griffiths and Gray 1994), and Uller and Helanterä (2019) proposed that heredity can be seen "*as [the] recurrence of developmental process*", which constitutes an excellent way to place development as a central evolutionary process.

Along the same line, Michael Skinner highlighted the major importance of early in life effects and underlined the necessity to incorporate development in interaction with epigenetics within the new synthesis (Skinner *et al.* 2010, Skinner 2011a, Skinner *et al.* 2011, Crews *et al.* 2014, Guerrero-Bosagna and Skinner 2014, Skinner *et al.* 2015). More generally, the full literature on early in life effects can be considered as supporting that major idea[1](for a review, see for instance the special issue on 'Developing differences: early-life effects and evolutionary medicine' and particularly its introductory article Kuijper *et al.* 2019).

Elisabeth Pennisi (2008) reporting on a small meeting held in Altenberg, Austria, under the lead of Massimo Pigliucci and Gerd Müller summarized the opinion of most participants in calling (not to say urging) for the full community of evolutionary biologists to 'Modernize the Modern Synthesis'. That idea clearly dominated the multi-author book the organizers of that meeting published two years later (Pigliucci and Müller 2010). That group insisted on the

---

[1] For a review, see for instance the special issue on 'Developing differences: early-life effects and evolutionary medicine' and particularly its introductory article Kuijper *et al.* 2019.

importance of epigenetic inheritance and development as major actors of evolution. They claimed that "*the Modern Synthesis needs to better incorporate modern science*" (Pennisi 2008). I could not say it better.

The next step, however, is to go beyond calling for a new synthesis and proposing a way to build it, or if you prefer sketching what it could look like. This was where evolutionary ecologists needed to get into the play.

18.2.5. Evolutionary ecologists finally got on-board: heredity is pluralistic

Only relatively recently, mainly since 2000, evolutionary ecologists took on-board the train of the modernisation of the Modern Synthesis. This delay probably reveals that the Modern Synthesis influenced far more evolutionary ecologists than any other biologists, because that paradigm constituted their everyday general framework. Therefore, it was probably much harder for evolutionary ecologists (i.e. those biologists specialized in the study of mechanisms of evolution) to contemplate the possibility that their long accepted general framework might be incomplete and might deserve some rethinking. We know that models are crude simplification of reality. However, if seeking simplification is a necessary step forward, we should not forget to distrust our simplifications as soon as it becomes necessary. As always in science, facts are stubborn and discoveries resulting from the advent of high-throughput sequencing forced some evolutionary ecologists to realize that the Modern Synthesis was too incomplete to accommodate the quickly growing evidence for non-genetic inheritance, indeed in other domains of biology, but that nonetheless reached them.

Evolutionary ecologists formalized the fact that the debate on the potential role of epigenetics and development in evolution concerns mechanisms of inheritance. *All the newly discovered processes naturally participate to heredity*, i.e. to resemblance that directly results from the influence of ancestors on the phenotype of their descendants. For instance, it is striking that Pigliucci and Müller (2010) placed "inheritance" at the centre of their Figure 1.1, thus revealing the key role they attribute to mechanisms of inheritance. For instance, the effect of plastic responses to a single environmental stress that have been shown to persist for at least 80 generations afterwards in *C. elegans* (Vastenhouw *et al.* 2006) would certainly be captured by any proper heritability estimation method of the corresponding trait leading to overestimate heritability in its genetic (i.e. sequencic) meaning. Applying the tools of quantitative genetics to integrate the evolutionary effects of all the mechanisms put forward by students of animal culture, epigeneticists and developmental biologists into a new evolutionary synthesis thus constitutes a promising avenue. Following the application of the logic of quantitative genetics to culture inheritance as in Danchin *et al.* (2004), Russell Bonduriansky (2012) proposed to rethink heredity. His point was that, after decades during which inheritance as the result of the sole gene transmission was the mainstream view, we now need to accept the pluralistic nature of inheritance. A few years before, Bonduriansky and his colleague Troy Day had proposed a new approach that "*shows that, by decoupling phenotypic change from the genotype, nongenetic inheritance can circumvent the limitations of genetic inheritance and thereby influence population dynamics and alter the fitness landscape.*" (Bonduriansky and Day 2009). Their modelling was based on a review of recent evidence from various fields of biology, including epigenetic, cytoplasmic and somatic, environmental and cultural inheritance. Their review, as well as a long series of review articles on the topic, supported the major idea that we need to consider the effects of the many newly documented processes as participating to inheritance alongside the transmission of sequencic information. Their generic framework derived from a Mendelian genetic model of inheritance decoupled phenotypic change resulting from pure genotypic change on the one hand, from the one brought about by non-genetic inheritance on the other hand, in order to consider the consequences for the study of evolution (Bonduriansky and Day 2009, Day and Bonduriansky 2011). They concludes that that decoupling constitutes "*a potent factor in evolution that can engender outcomes unanticipated under the Mendelian-genetic model*" (Bonduriansky and Day 2009).

Bonduriansky and Day (2009) also tackled the question of the possibility of the inheritance of acquired phenotypes, a taboo within the Modern Synthesis at the time. Today, the accumulation of evidence for mechanisms of transmission of acquired phenotypes (such as depression, fears, or diabetes among many other examples) over many generations demonstrates the existence of such form of inheritance in many organisms on many traits, and appears to result from sophisticated processes of soma-to-germen communication (among many reviews see for instance Bonduriansky and Day 2009, Danchin *et al.* 2011, Wang *et al.* 2017, Bonduriansky and Day 2018, Danchin *et al.* 2019b). The longest-lasting documented example of inheritance of acquired character that I know of was reported to last at least 80 generations in *C. elegans* (Vastenhouw *et al.* 2006), an amazing durability leaving plenty of time for natural selection to act on such highly heritable characteristics.

Still along the idea of using methodologies developed in the context of the Modern Synthesis, Helanterä and Uller (2010) adapted the price equation to explore the non-consensual question of the evolutionary consequences of the four types of recently identified inheritance systems by Jablonka and Lamb (2005). They conclude that although the four inheritance systems can share conceptually very similar features, their implications for adaptive evolution can differ substantially because of differences in their ability to couple natural selection and inheritance. More recently, the same authors explored the theoretical implications of the existence of various forms of non-genetic inheritance regarding evolutionary dynamics (Uller and Helanterä 2019). More specifically, they analyse whether it is necessary to build a new synthesis that would integrate the recent discoveries on the non-genetic inheritance of traits, or whether the Modern Synthesis can accommodate those new discoveries without any significant change. This remains on open question.

18.2.6. The above review is incomplete

The above review is far from complete, as I did not develop cytoplasmic inheritance, ecological inheritance, niche construction, inheritance resulting from the transmission of prions, chaperon molecules or the microbiota (Pigliucci 2005, Shorter and Lindquist 2005, Halfmann and Lindquist 2010, Danchin *et al.* 2011, Lindquist 2011, See for instance: Halfmann *et al.* 2012, Manjrekar 2017, Muller 2017, Newby *et al.* 2017, Bonduriansky and Day 2018, reviews in Danchin *et al.* 2019b, Uller and Helanterä 2019). These are domains that need to be integrated too as there is evidence that their influence on parent-offspring resemblance is probably far from negligible.

18.3Towards the new evolutionary synthesis

aa Undoubtedly, the sequencic simplification of the Modern Synthesis was a necessary step in the 1950 and the following decades, as science always progresses by simplification as a first step. The risk, however, is that after some time, the reduced version may become the truth, as researchers tend to forget that the currently mainstream paradigm was intended initially to be a temporary simplification. This attitude then makes it very hard to overcome the simplification to cross another scientific step that will necessarily be another simplification, but one that accepts a little bit more complexity. This is exactly what the paradigm of evolution is undergoing now, and it is the ultimate aim of this book to accompany the evolutionary community in taking one of these steps from the Modern Synthesis to a new synthesis.

My goal in this section is to discuss the level of ambition that we should adopt in that step. Should we just extend the Modern Synthesis by adding the fantastic developments brought about by the recent discoveries in the domain of epigenetics, or should we be more ambitious in also integrating all other forms of non-genetic (i.e. non-sequencic) inheritance?

18.3.1 The Extended Evolutionary Synthesis

The first phrase that emerged to name the new synthesis of evolution is the “Extended Evolutionary Synthesis”. Since many of the initial proponents of the extended synthesis come from functional biology (in the meaning of Mayr 1961) and particularly molecular biology, the emphasis is mostly on the importance of development and epigenetics. According to them, the new synthesis must incorporate the fantastic memory system of epigenetics. I fully agree that this is a necessary step forward.

However, although important, this can only be an intermediate step towards the new evolutionary synthesis for the 21rst century, as the new synthesis should integrate *all* the pathways of intergenerational transfer of information involved in resemblance among relatives and interacting non-relatives. One of the obvious consequences of the sequencic flavour of the Modern Synthesis in its present form is that many current functional biologists (again in the meaning of Mayr) tend to consider that only approaches involving molecular biology really "advance science". The tendency to just integrate epigenetics is all the stronger since the approaches classically used to study non-genetic inheritance beyond epigenetics do not necessarily involve molecular approaches (at least initially). This is the case for instance for cultural and ecological inheritance or niche construction. All this suggests that there is a great risk that the extended synthesis will only incorporate epigenetics in its classical sense.

A recent example illustrates this risk. I was contacted by one of the authors of an article that had been just published in *Trends in Ecology and Evolution* and who, knowing my work, wanted to inform me of the publication of that article (Adrian-Kalchhauser *et al.* 2020). The title of their article "*Understanding 'Non-genetic' Inheritance: Insights from Molecular-Evolutionary Crosstalk*" seemed excellent. However, the article used formulations that repeatedly implied that non-genetic inheritance can be reduced to what they call "*inherited gene* *regulation*", which corresponds to one of the possible definitions of epigenetics (Danchin 2022). Insidiously, this conveys the message that non-genetic inheritance boils down to its epigenetic component, thus forgetting all the other dimensions of inclusive inheritance, namely the transmission of cultural, ecological, prions, chaperone molecules and cytoplasmic or microbiota variants, as well as niche construction. Although rarely responding to articles with which I disagree, having been contacted by *Pim Edelaar* from the University *Pablo de Olavide* in Seville in Spain, I nonetheless agreed to participate to a response as I felt the stakes were too high to remain silent. The main message of our response (Edelaar *et al.* 2021) is that this view of the new synthesis amounts to repeating the reductionist error of the Modern Synthesis, since we now clearly know that non-genetic inheritance cannot be reduced to the mere inheritance of genes and gene regulation. It should be noted, however, that the Modern Synthesis did not really make such a mistake when it was established because, at the time, sequencic was the only known pathway of inheritance. Contrastingly, today, only adding epigenetics would make this mistake because we know many other mechanisms of parent-offspring resemblance. Interestingly, the authors replied to our comment saying that their objective was definitely not to reduce non-genetic inheritance to inherited gene regulation, and that therefore we fully agree (Adrian-Kalchhauser *et al.* 2021). Thus, if you think about it, the fact that the intention of these authors was not reductionist shows how insidiously the reductionist approach can creep through all our reasoning. Without us even realising it, that attitude influences even the most open-minded people like these authors.

More generally, considering that only the study of infra-individual processes (i.e. Mayr's functional biology) is relevant to biology, amounts to rejecting the nonetheless largely accepted claim of Tinbergen (1963) that all four approaches (proximate, development, fitness effect and macro evolution) to evolution are legitimate and *jointly* necessary to understand evolution.

18.3.2 The Inclusive Evolutionary Synthesis

In fact, the attitude favouring proximate approaches (i.e. Tinbergen's approaches on proximate mechanisms and development, which together correspond to Mayr's functional biology) is unfortunately so common among functional biologists that it constitutes one of the major brakes for the emergence of a more ambitious new synthesis. In my opinion, this justifies giving the new synthesis another name. This is why I have been calling since 2010 and especially since 2013, when I first published the expression ‘Inclusive Evolutionary Synthesis’ (IES), for the new evolutionary synthesis to be *Inclusive* rather than just *Extended*.

What, in my opinion, profoundly differentiates these two approaches is a question of ambition. As I said above, the Extended Evolutionary Synthesis runs the risk of adding only the epigenetic dimension. This would integrate all the properties of the DNA molecule, which would be a wonderful and necessary step. But it is not enough. Our ambition must be to integrate not only epigenetics, but also all the other mechanisms of resemblance that emerge from the stable and transmissible properties of other major molecules (such as proteins) or that emerge from stable and transmitted states at higher levels of organisation of living beings (tissues, organs or even complete organisms as with behaviour). This ambition of comprehensiveness justifies the qualifier ‘Inclusive’, rather than just ‘Extended’, in order to incorporate the effects of all forms of transmission that can play a role in inheritance and evolution.

Ignoring mechanisms of resemblance resulting from the properties of other molecules than the DNA, or emerging at higher levels of organisation in fact boils down to denying the existence of emergent properties (Anderson 1972, Salt 1979), which are properties of a given entity that are more than the sum of the properties of its components. This implies that the study of the properties of an entity cannot be deduced from the sole use of the concepts and tools developed to study its components. A striking example of emergent property is that of the brain (an organ) whose functioning produces thoughts. Thoughts emerge from the complex interactions among the circa one hundred billions cells of the brain, and nobody would pretend that we can understand thoughts by only using the classical tools and concepts of neurobiology (e.g. electrodes, spikes, calcium pumps etc). To understand thoughts, we need to use disciplines of a very different nature, such as psychology, mathematics, and philosophy. My point is that claiming that all the properties of living organisms could be summed up in the sole properties of the DNA molecule, whose sequencic and 3D structure would allow us to fully understand the complexity of living organisms in fact boils down to ignoring the many important properties of living entities that emerge at levels of organisation higher than that of the DNA (for more discussion about the links between the 3D or 4D structure of the DNA and epigenetics, see Danchin 2022).

Furthermore, the concept of emergent property is at the heart of the major evolutionary transitions proposed by Maynard Smith and Szathmáry (1995), Szathmàry and Maynard Smith

(1995). Hence, limiting the new synthesis to the addition of the role of epigenetics in inheritance would, in effect, only incorporate the first three major transitions into the new synthesis, namely (*i*) the transition from replicating molecules to populations of molecules into compartment, (*ii*) from independent replicators to chromosomes and (*iii*) from RNA as gene and enzyme to DNA plus protein (i.e. the genetic code). In effect, only incorporating epigenetic inheritance would ignore the five other major transitions identified by these authors. In fact, most of the documented mechanisms of non-genetic inheritance are consubstantial with most of the eight major transitions they identified and particularly so with the five last transitions, namely (*iv*) from pro- to eukaryotes, (*v*) from asexual clones to sexual populations, (*vi*) from protists to multicellular organisms, (*vii*) from solitary individuals to sociality (eusociality), and (*viii*) from primate societies to human language.

An interesting anecdote, however, is that I very recently discovered that I was not the first to use the phrase Inclusive Evolutionary Synthesis to qualify the long sought new synthesis as I found it in the title of Sommer (2004), which was a review of Wallace Arthur's very interesting book (Arthur 2004), nine years before I first used it in a paper. In that book, Wallace Arthur advocates the necessity to incorporate the major role of development into the new synthesis. As he recurrently uses the phrase "inclusive synthesis" in his Chapter 16, he should be granted for coining that term (although he never used that exact expression). Obviously (and fortunately), I am not the only one to consider that inclusiveness should be a major characteristic of the new evolutionary synthesis. Interestingly, today, Arthur's proposal to integrate development would be no more inclusive as in the meantime a plethora of new processes of inheritance were discovered.

To sum up, I suggest that we should consider the *Extended* Evolutionary Synthesis as an intermediate step in the ongoing emergence of a new synthesis. The term "*inclusive*" in the phrase "Inclusive Evolutionary Synthesis" flags the general ambition to be broader in integrating all known mechanisms of transmitted resemblance into the evolutionary synthesis that we are building for the 21st century.

18.3.3 My implication in the emergence of the new synthesis

I perfectly remember the first time I was confronted to the idea that inheritance may not boil down to the transmission of the DNA sequence. It was in 2000, while writing the chapter about sexual selection for the textbook on Behavioural Ecology that we published in French in 2005 (Danchin *et al.* 2005) and then in English in 2008 (Danchin *et al.* 2008b). I was writing that chapter with Franc Cézilly, and although he did not believe in it, he said incidentally that we could have a section about cultural inheritance in the context of sexual selection. Honestly, I did not understand what he was talking about. For me, there was only one kind of inheritance, and it was genetic (in the meaning of sequencic). I was a perfect tenant of the Modern Synthesis of Evolution, to which "The gene's eye view" of biology as developed by J. Arvid Ågren (2021) constitutes one important component.

Without realizing it, Frank had instilled doubt in my head and I started reading about animal culture. I was so fascinated by what I discovered, that I added two new chapters on the issue to the English version, one about social learning (about which I already had some knowledge Danchin *et al.* 2008a) and one about cultural inheritance (Danchin and Wagner 2008). It was followed by the publication of a review article (Danchin *et al.* 2004) in which we developed major ideas about the use of public information (i.e. information that is accessible to all, as opposed to personal information) as its potential to lead to cultural transmission and used a quantitative genetics approach to define culture.

Since then, I have kept claiming that all the arguments and experimental evidence related to the role of epigenetics, development, animal culture, the transmission of environments, cell types, microbiota, prions and chaperon molecules, as well as niche construction in fact directly affect heredity at the scale of individuals within populations. My point is that, although it is a necessary step, we should not be satisfied with the sole detailed description of these fascinating mechanisms. We must also go at least one-step farther by integrating them into a broader concept of heredity (i.e. patterns of transmitted resemblance) and inheritance (i.e. the study of the various mechanisms that produce such patterns of resemblance). In fact, together heredity and inheritance constitute one of the keystones of the conceptual edifice of the new synthesis. This is because they are ideal concepts to connect what I call "infra-individual biology" (and that Ernst Mayr (1961) called "Functional Biology") with what I call "supra-individual biology" (corresponding to Ernst Mayr (1961) "evolutionary biology" (see for instance: Danchin and Pocheville 2014)).

18.3.4 The biggest challenge for the new synthesis

In effect, heredity certainly constitutes a real bridge between these two vast domains of biology that have been separated by all our institutions for too long so that now, not only do they usually ignore each other, but also often despise each other. Historically, the ancient separation between functional and evolutionary biology naturally resulted from the fact that these two domains focus on very different levels of organisation of living entities, and thus necessitate very different methodologies, tools and concepts. Molecules constitute the most basic entity of functional biology, which encompasses anatomy, genetics in its molecular sense, cell biology and physiology (with its suite of molecular approaches and sub-domains). Contrastingly, individuals constitute the most basic entity of evolutionary biology that encompasses behaviour, population approaches (demography and quantitative genetics), ecology (interactions among individuals and species within communities and ecosystems), biodiversity and the biological part of climate studies.

These profound differences have long led most of our institutions to separate these disciplines in clearly distinct departments. The result is that today these two domains constitute independent worlds that do not communicate, so that to me, bringing the whole biology to work together constitutes the most difficult challenge to establish the new synthesis that will at last reunify the whole of biology into a single integrative framework. This challenge is made more complicated by the use of terms like "system biology" or "integrative biology" only within functional biology. These terms are thus loaded, while they would be perfectly suitable to the far more integrative approach of merging functional biology with evolutionary biology (again in Mayr's meaning) as this should be the goal of the new synthesis.

I personally experienced this challenge while leading a large-scale project during the last 15 years of my carrier. In 2006, my lab entitled "Évolution & Diversité Biologique" joined other labs working on plant biology (plant development, plant symbiosis, plant pathology etc.) within a consortium of labs. In 2010, that consortium applied to a French state call for proposal in order to transform the consortium into a multi-millions euros project with 5-year terms. Dominique Roby, Jean Clobert, Jacques Batut and I wrote the project, which got funded under the name TULIP (Towards a unified theory of biotic interactions: role of environmental perturbations). Dominique Roby and I took the lead of TULIP for two terms, at the end of which we obtained a second renewal for another 5-year term. Its essence was to lead people working in functional and evolutionary biology to work together, and one of its implicit ideas was to test what I was starting to call the Inclusive Evolutionary Synthesis. Initially, TULIP encompassed five labs involving about 400 staff (researchers, university teachers, administrators, and PhD students). In 2019, founding labs had grown in size and had been joined by two other labs leading to circa 700 staff. In parallel, we managed to raise funds for a new building for some offices and labs of the two concerned scientific domains. Finally, at the end of our second term, TULIP fostered the emergence of a TULIP graduate school in order to train our students within the TULIP philosophy. We were thus quite successful in relation to our institutions, but internally, we experienced the many difficulties of transdisciplinary challenges. Both the functional and evolutionary halves were excellent (which is a prerequisite for the success of a merging endeavour). We increased the number, quality and impact factor of the articles we published; the number of papers having authors belonging to both disciplines increased steadily during the ten years, though they remained rare and represented a minute proportion of our total production.

The biggest challenge lied in the mentality of our interactions within the transdisciplinary project. Many speakers never fathomed that the audience was heterogeneous, encompassing members from a constellation of disciplines. Hence, many seminars remained obscure to a significant fraction of the audience. Even me, in spite of my conviction of the necessity to integrate our disciplines, and after 15 years of constant effort, I regularly got lost after the first few slides of talks on plant sciences. I even had the impression that there was some snobbism in trying to lose a good deal of the audience. I have been confronted to such incapacity to communicate with members of other disciplines every day of the ten years of my co-leading of the TULIP project. A recurrent issue is that each discipline is convinced that its vision of causality is far better than that of other disciplines (a topic already tackled by Mayr 1961). This is one of the Gordian knot of transdisciplinarily. I cannot recall how many times I heard the phrase “this is only descriptive”, while I regularly felt “why are they doing those studies? In other words, what is their scientific question?”, or “yes, but you still haven’t demonstrated causality, you are still stuck in some sort of correlations”.

However, this long experience has had positive sides. For instance, the initial language problems were relatively easy to circumvent. Also, the best scientific moments for me were our yearly international summer schools. We invited top scientists from all over the world to give lectures in the morning and insisted for them to be pedagogical as the audience was by definition quite heterogeneous. Although all the speakers were stars in their domain, this is the only moment in which I felt that I understood the messages of functional biologists, even towards the end of their talks when they accelerated in tackling the current challenges of their disciplines. In these enlightening moments, I regularly thought that we had a lot to do together, but I seldom felt this after our internal seminars.

My point in describing my own experience is to show an overlooked key issue for the success of transdisciplinary projects. *More than ever, we must work on how we talk about our research within the project.* We should not ambition to convince the audience that we are at the front of our discipline; this is not the point. On the contrary, we should put a lot of energy on the pedagogy of our talks and discussions. This is the only way to get understood and to leave all doors and all borders open.

It was clear to me, right from the beginning that the challenge of unifying disciplines separated for generations would be a multi-decadal challenge, but god, that was a slow motion! Again, to me the only true challenge for the real emergence of a new integrative synthesis is not in the purely scientific range but rather in the psychological capacity of members from various research areas to *listen to and respect* each other, the first step to envision integration.

My other experience of this kind is in my 10-year still ongoing collaboration with Guillaume Isabel, a neurobiologist of memory in fruit flies, in the study of the cultural transmission of mating preferences. It has been a highly fruitful collaboration because, right from the beginning, we have respected each other and have been willing to work hard to understand what the knowledge of the other could bring into our integrative study. Guillaume and I regularly state, in a simple language, the principle of this or that technology, or the rationale of that or this question etc. and this is very useful.

To sum up, I think that the shift from the Modern Synthesis to the new synthesis is a major step for our understanding of life, and slowness is not stagnation. Any exponential growth starts very slow, and nonetheless, after a while it looks more like an explosion. Thus, early slowness in the building of a new synthesis should not refrain us from acting in that direction.

18.3.5 A parallel with the conceptual revolution of relativity

This ongoing evolution (some say revolution) in biology from the Modern Synthesis of Evolution to the Inclusive Evolutionary Synthesis via what I consider as an intermediate step in the form of the Extended Evolutionary Synthesis is reminiscent of the transition from Newtonian physics to relativity, first special and then generalized at the beginning of the 20th century. I am perfectly conscious that such a parallel may seem grandiloquent, but nonetheless, similarities are striking, so that the parallel is intriguing enough to deserve consideration.

First and foremost, the transition from Newtonian physics to relativity did not invalidate Newtonian physics but generalized it to a vastly larger set of conditions and scales. In the same way, and contrary to the recurrent assertion that the discoveries made during the last decades invalidate the Modern Synthesis, the transition from the Modern Synthesis to the new synthesis (whether you call it extended or inclusive) does NOT invalidate the Modern Synthesis. Indeed, the recent discoveries merely scratch the surface of the conceptual edifice of the Modern Synthesis in questioning some of its basic principles. For instance, while the Modern Synthesis rejects the possibility of inheritance of acquired characters, the new synthesis should incorporate that possibility. Similarly, we should get rid of the oversimplification of inheritance that emerged in the second half of the 20th century concerning mechanisms of inheritance in accepting that parent-offspring resemblance does not result *only* from the transmission of the DNA sequence, but also includes a series of sophisticated mechanisms that we are just starting to uncover. Similarly, as developed in Danchin *et al.* (2019b), the important principle that *mutations are never oriented by the environment towards improving organisms' adaptation to it* remains entirely valid. What is incorrect, however, is the oversimplification of that principle often used within the Modern Synthesis that "mutations are random" (sometimes adding, "relative to the environment"). This last formulation is a caricature of the general principle, to which it constitutes a strawman[2]. In fact, I want to highlight here that the new synthesis is built on the Modern Synthesis and that the latter's many successes constitute the solid foundation on which to elaborate the new conceptual edifice.

---

[2] On randomness of mutations, see among others (Merlin 2010).

The second major similarity with the Einsteinian revolution in astrophysics, lies in the fact that both the early 20th century revolution, and the early 21rst century (r)evolution in progress in evolutionary biology unfolded in two steps. This similarity is particularly striking, as if both were too big a change to unfold in only one step. Another potential similarity that I anticipate stems from the fact that today we scarcely use special relativity, because it was an imperfect temporary step towards generalised relativity that is the only one now recognised as effective. I anticipate that the same will occur concerning the Extended *versus* Inclusive Evolutionary Synthesis. This is particularly important for instance in medicine, in which only accounting for the sequencic plus epigenetic components of inheritance would soon show its limitations when searching for new therapies.

However, the similarities between these two conceptual shifts stop there. For instance, although general relativity is a far better descriptor of the way the universe functions at all scales, agencies willing to send a rocket to the moon do not use relativity because, at that scale, Newtonian physics is sufficient. This is where the difference lies, as I do not think that the same holds with the new evolutionary synthesis as there is probably no examples of inheritance that only results from the transmission of pure sequencic information. This means that the sequencic view of inheritance that dominated the Modern Synthesis for 70 years is likely to show its limits, whatever the kind of trait under study. Currently, limitations are showing off in every corner of medical sciences for instance, as with the case of missing heritability. The Modern Synthesis has now reached its limitations and we need to proceed forward towards a new synthesis that has more predictive and practical power, particularly in medicine and conservation biology.

Another question on the new evolutionary synthesis concern whether it will significantly change the way we do science, in other words whether the corresponding paradigm shift corresponds to what *Thomas Kuhn* called "extraordinary science", which marks the passage from one "normal science" to another (Kuhn 2021). However, on top of the fact that we would need far more information, and particularly a lot of modelling, before answering such a question, this is certainly not the role of tenants of a new idea to decide whether it consists in a paradigm shift in Kuhn's meaning or not. This is the role of historians of science to decide, and I'm afraid we'll have to wait for a few decades before getting an answer on this issue.

As I explained here, we should work hard to ensure that the new synthesis does not forget about non-genetic processes of inheritance beyond the incorporation of epigenetics. To me, it is the role of evolutionary biologists and in particular behavioural ecologists to be convincing enough to lead the whole of biology to incorporate all the inheritance systems beyond those concerning the DNA sequence and 3D structure. As I explain, this is the only way to incorporate into the new synthesis the many kinds of emergent properties that piled up one on top of the other, step by step, during the 3.5 billion years of evolution on planet Earth. And this is also the only way to integrate all the major evolutionary transitions into the new synthesis. There are many reasons why the new synthesis might not be inclusive, and these are so pregnant in present day biology that they constitute the main difficulty to overcome to establish a comprehensive new evolutionary synthesis. This is why I wanted to anticipate this trap and try to influence the course of evolutionary sciences by coining a name for the new synthesis that clearly flags its ambition of completeness or inclusivity. My deep hope is that the term inclusive will remind any scientist that the sole incorporation of epigenetics is insufficient. Only the future will tell whether this ambition has been endorsed by the bulk of the community of evolutionary biologists. This is my dearest wish.

Acknowledgements

I thank Agathe du Crest for having carefully commented and edited this manuscript, hence significantly improving the chapter. My studies were supported by the CNRS, the Université Paul Sabatier, and the "Laboratoires d’Excellence (LABEX)" TULIP (ANR-10-LABX-41).

References

Adrian-Kalchhauser I, Sultan SE, Shama LNS, Spence-Jones H, Tiso S, Valsecchi CIK, Weissing FJ (2020) Understanding'Non-genetic'Inheritance: Insights from Molecular-Evolutionary Crosstalk. **Trends Ecol Evol** 35:1078-1089

Adrian-Kalchhauser I, Sultan SE, Shama LNS, Spence-Jones H, Tiso S, Valsecchi CIK, Weissing FJ (2021) Inherited Gene Regulation Unifies Molecular Approaches to Nongenetic Inheritance: Response to Edelaar et al. **Trends Ecol Evol** 36:477

Ågren JA (2021) The Gene's-Eye View of Evolution. Oxford University Press, Oxford

Alem S, Perry CJ, Zhu X, Loukola OJ, Ingraham T, Søvik E, Chittka L (2016) Associative Mechanisms Allow for Social Learning and Cultural Transmission of String Pulling in an Insect. **PloS Biol** 14:e1002564

Allen J, Weinrich M, Hoppitt W, Rendell L (2013) Network-Based Diffusion Analysis Reveals Cultural Transmission of Lobtail Feeding in Humpback Whales. **Science** 340:485-488

Anderson PW (1972) More is different. **Science** 177:393-396

Anway MD, Cupp AS, Uzumcu M, Skinner MK (2005) Epigenetic transgenerational actions of endocrine disruptors and mate fertility. **Science** 308:1466-1469

Aplin LM, Farine DR, Morand-Ferron J, Cockburn A, Thornton A, Sheldon BC (2015) Experimentally induced innovations lead to persistent culture via conformity in wild birds. **Nature** 518:538-541

Arthur W (2004) Biased embryos and evolution. Cambridge University Press, Cambridge

Avital E, Jablonka E (2000) Animal Traditions. Behavioural Inheritance in Evolution. Cambridge University Press, Cambridge

Bohacek J, Mansuy IM (2015) Molecular insights into transgenerational non-genetic inheritance of acquired behaviours. **Nature Rev Genet** 16:641-652

Bonduriansky R (2012) Rethinking heredity, again. **Trends Ecol Evol** 27:330-336

Bonduriansky R, Day T (2009) Nongenetic inheritance and its evolutionary implications. **Annu Rev Ecol Syst** 40:103-125

Bonduriansky R, Day T (2018) Extended Heredity: A New Understanding of Inheritance and Evolution. Princeton University Press, Princeton and Oxford

Bossdorf O, Richards CL, Pigliucci M (2008) Epigenetics for ecologists. **Ecol Let** 11:106-115

Bourrat P, Lu Q, Jablonka E (2017) Why the missing heritability might not be in the DNA. **BioEssays** 39:1700067

Bourrat P, Lu QY (2017) Dissolving the Missing Heritability Problem. **Phil Sci** 84:1055-1067

Boyd R, Richerson PJ (1983) Why is culture adaptive? **Q Rev Biol** 58:209-214

Boyd R, Richerson PJ (1985) Culture and the evolutionary process. Chicago University Press, Chicago

Boyd R, Richerson PJ (1995) Why does culture increase human adaptability. **Ethol Sociobiol** 16:125-143

Brown CR, Laland KN (2003) Social learning in fishes: a review. **Fish and Fisheries** 4:280-288

Cavalli-Sforza LL, Feldman MW (1981) Cultural transmission and evolution: a quantitative approach. Princeton University Press, Princeton NJ

Cavalli-Sforza LL, Feldman MW (1983) Cultural versus genetic adaptation. **Proc Nat Acad Sci USA** 80:4993-4996

Champagne FA (2008) Epigenetic mechanisms and the transgenerational effects of maternal care. **Front Neuroendocrinol** 29:386-397

Champagne FA, Curley JP (2009) Epigenetic mechanisms mediating the long-term effects of maternal care on development. **Neurosci Biobehav Rev** 33:593–600

Chapman BB, Ward AJW, Krause J (2008) Schooling and learning: early social environment predicts social learning ability in the guppy, Poecilia reticulata. **Anim Behav** 76:923-929

Coolen I, Dangles O, Casas J (2005) Social learning in noncolonial insects? **Curr Biol** 15:1931-1935

Crews D, Gillette R, Miller-Crews I, Gore AC, Skinner MK (2014) Nature, nurture and epigenetics. **Mol Cell Endocrinol** 398:42-52

Crews D, Gore AC, Hsu TS, Dangleben NL, Spinetta M, Schallert T, Anway MD, Skinner MK (2007) Transgenerational epigenetic imprints on mate preference. **Proc Nat Acad Sci USA** 104:5942-5946

Danchin É (2013) Avatars of information: towards an inclusive evolutionary synthesis. **Trends Ecol Evol** 28:351-358

Danchin É (2022) More than fifty shades of Epigenetics for the study of early in life effects in medicine, ecology and evolution. In: Costantini D, Marasco V (eds) Development Strategies and Biodiversity - Darwinian Fitness and Evolution in the Anthropocene. Springer Nature Switzerland AG

Danchin É, Boulinier T, Massot M (1998) Conspecific reproductive success and breeding habitat selection: implications for the study of coloniality. **Ecology** 79:2415-2428

Danchin É, Charmantier A, Champagne FA, Mesoudi A, Pujol B, Blanchet S (2011) Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. **Nature Rev Genet** 12:475-486

Danchin É, Giraldeau L-A, Wagner RH (2008a) An information driven approach to behaviour. In: Danchin É, Giraldeau L-A, Cézilly F (eds) Behavioral Ecology. Oxford University Press, pp. 97-131

Danchin É, Giraldeau LA, Cézilly F (2005) Ecologie comportementale. Dunod, Paris

Danchin É, Giraldeau LA, Cézilly F (2008b) Behavioural Ecology. Oxford University Press, London

Danchin É, Giraldeau LA, Valone TJ, Wagner RH (2004) Public information: from nosy neighbors to cultural evolution. **Science** 305:487-491

Danchin E, Isabel G, Nöbel S (2022) Culture in insects. In: Tehrani J, Kendal R, Kendal J (eds) Oxford Handbook of cultural evolution.

Danchin É, Nöbel S, Pocheville A, Dagaeff A-C, Demay L, Alphand M, Ranty-Roby S, van Renssen L, Monier M, Gazagne É, Allain M, Isabel G (2018) Cultural flies: conformist social learning in fruit flies predicts long-lasting mate-choice traditions. **Science** 362:1025-1030

Danchin É, Pocheville A (2014) Inheritance is where physiology meets evolution. **J Physiol** 592:2307-2317

Danchin É, Pocheville A, Huneman P (2019a) Early in life effects and heredity: reconciling neo-Darwinism with neo-Lamarckism under the banner of the inclusive evolutionary synthesis. **Phil Trans R Soc B** 374:20180113

Danchin É, Pocheville A, Rey O, Pujol B, Blanchet S (2019b) Epigenetically-Facilitated mutational assimilation: Epigenetics as a hub within the Inclusive Evolutionary Synthesis. **Biol Rev** 94:259-282

Danchin É, Pujol B, Wagner RH (2013) The double pedigree: a method for studying culturally and genetically inherited behavior in tandem. **PLoS ONE** 8:e61254

Danchin É, Wagner RH (1999) Commodity selection and coloniality: Exagerated traits are produced by processes of choice. In: Adams N, Slotow R (eds) Proceedings of the 22nd International Ornithological Congress. University of Natal Press

Danchin É, Wagner RH (2008) Cultural evolution. In: Danchin É, Giraldeau L-A, Cézilly F (eds) Behavioral Ecology. Oxford University Press, pp. 693-726

Danchin É, Wagner RH (2010) Inclusive heritability: combining genetic and nongenetic information to study animal behavior and culture. **Oikos** 119:210-218

Dawkins R (1976) The selfish Gene. Oxford University Press, Oxford

Dawkins R (1982) The Extended Phenotype. The Gene as the Unit of Selection. Oxford University Press, New York

Day T, Bonduriansky R (2011) A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. **Am Nat** 178:E18-E36

Dias BG, Ressler KJ (2014) Parental olfactory experience influences behavior and neural structure in subsequent generations. **Nature Neurosci** 17:89-96

Dobzhansky T (1973) "Nothing in Biology Makes Sense Except in the Light of Evolution". 35:125-129

Doligez B, Cadet C, Danchin É, Boulinier T (2003) When to use public information for breeding habitat selection? The role of environmental predictability and density dependence. **Anim Behav** 66:973-988

Doligez B, Danchin É, Clobert J (2002) Public information and breeding habitat selection in a wild bird population. **Science** 297:1168-1170

Doligez B, Danchin É, Clobert J, Gustafsson L (1999) The use of conspecific reproductive success for breeding habitat selectionin a non-colonial, hole-nesting species, the collared flycatcher. **J Anim Ecol** 68:1193-1206

Edelaar P, Bonduriansky R, Charmantier A, Danchin É, Pujol B (2021) Response to Kalchhauser et al.: inherited gene regulation is not enough to understand non-genetic inheritance. **Trends Ecol Evol** 36:475-476

Feldman MW, Cavalli-Sforza LL (1984) Cultural and biological evolutionary processes: gene-culture disequilibrium. **Proc Nat Acad Sci USA** 81:1604-1607

Feldman MW, Cavalli-Sforza LL (1989) On the theory of evolution under genetic and cultural transmission with application to the lactose absorption problem. In: Feldman MW (eds) Mathematical evolutionary theory. Princeton University Press

Fisher RA (1930) The genetical theory of natural selection. Clarendon Press, Oxford

Francis D, Diorio J, Liu D, Meaney MJ (1999) Nongenomic transmission across generations of maternal behavior and stress responses in the rat. **Science** 286:1155-1158

Franklin TB, Mansuy IM (2010) Epigenetic inheritance in mammals: Evidence for the impact of adverse environmental effects. **Neurobiol Disease** 39:61–65

Galef BG (1990) A historical perspective on recent studies of social learning about foods by Norway rats. **Can J Psychol** 44:311-329

Griffin AS (2004) Social learning about predators: A review and prospectus. **Learnin Behav** 32:131-140

Griffiths PE, Gray RD (1994) Developmental systems and evolutionary explanation. **J Philosoph** 91:277-304

Guerrero-Bosagna C, Skinner MK (2012) Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. **Mol Cell Endocrinol** 354:3-8

Guerrero-Bosagna CM, Skinner MK (2014) Environmental epigenetics and phytoestrogen/phytochemical exposures. **The J Steroid Biochem Mol Biol** 139:270-276

Haldane JB (1932 [1990]) The causes of evolution. Princeton University Press,

Halfmann R, Jarosz DF, Jones SK, Chang A, Lancaster AK, Lindquist S (2012) Prions are a common mechanism for phenotypic inheritance in wild yeasts. **Nature** 482:363-368

Halfmann R, Lindquist S (2010) Epigenetics in the Extreme: Prions and the Inheritance of Environmentally Acquired Traits. **Science** 330:629-632

Hamilton WD (1964) The Genetical Evolution of Social Behaviour I. and II. **J Theor Biol** 7:1-16 and 17-52

Helanterä H, Uller T (2010) The Price equation and extended inheritance. **Phil Theo Biol** 2:1-17

Heyes CM (1994) Social learning in animals: categories and mechanisms. **Biol Rev** 69:207-231

Heyes CM, Galef BGJ (1996) Social learning and Imitation: the Roots of Culture. Academic Press, New York

Jablonka E (2013) Epigenetic inheritance and plasticity: the responsive germline. **Progr Biophys Mol Biol** 111:99–107

Jablonka E, Lamb MJ (1989) The inheritance of acquired epigenetic variations. **J Theor Biol** 139:69-83

Jablonka E, Lamb MJ (1995) Epigenetic inheritance and evolution: The Lamarckian case. Oxford University Press, Oxford

Jablonka E, Lamb MJ (2005) Evolution in four dimensions. Genetic, Epigenetic, Behavioural, and Symbolic Variation in the history of life. MIT Press, Cambridge Massachusetts

Jablonka E, Lamb MJ (2010) Transgenerational epigenetic inheritance. In: Pigliucci M, Müller GB (eds) Evolution: The extended synthesis. MIT Press, pp. 137-174

Jablonka E, Lamb MJ, Avital E (1998) 'Lamarckian' mechanisms in darwinian evolution. **Trends Ecol Evol** 13:206-210

Jablonka E, Oborny B, Molnar I, Kisdi E, Hofbauer J, Czaran T (1995) The adaptive advantage of phenotypic memory in changing environments. **Phil Trans R Soc B** 350:133-141

Jesmer BR, Merkle JA, Goheen JR, Aikens EO, Beck JL, Courtemanch AB, Hurley MA, McWhirter DE, Miyasaki HM, Monteith KL (2018) Is ungulate migration culturally transmitted? Evidence of social learning from translocated animals. **Science** 361:1023-1025

Krützen M, Kreicker S, MacLeod CD, Learmonth J, Kopps AM, Walsham P, Allen SJ (2014) Cultural transmission of tool use by Indo-Pacific bottlenose dolphins (Tursiops sp.) provides access to a novel foraging niche. **Proc R Soc Lond B Biol Sci** 281:20140374

Kuhn T (2021) The structure of scientific revolutions. Princeton University Press,

Kuijper B, Hanson MA, Vitikainen EIK, Marshall HH, Ozanne SE, Cant MA (2019) Developing differences: early-life effects and evolutionary medicine. **Phil Trans R Soc B** 374:20190039

Lachmann M, Jablonka E (1996) The inheritance of phenotypes: An adaptation to fluctuating environments. **J Theor Biol** 181:1-9

Laland KN (1994) Sexual selection with a culturally transmitted mating preference. **Theor Pop Biol** 45:1-15

Laland KN, Odling-Smee J, Myles S (2010) How culture shaped the human genome: Bringing genetics and the human sciences together. **Nature Rev Genet** 11:137-148

Leadbeater E, Chittka L (2007) Social learning in insects - From miniature brains to consensus building. **Curr Biol** 17:R703-R713

Leadbeater E, Raine NE, Chittka L (2006) Social learning: Ants and the meaning of teaching. **Curr Biol** 16:R323-R325

López-Cortegano E, Caballero A (2019) Inferring the nature of missing heritability in human traits using data from the GWAS catalog. **Genetics** 212:891-904

Lu Q, Bourrat P (2017) The evolutionary gene and the extended evolutionary synthesis. **British J Phil Sci** 69:775-800

Maher B (2008) Personal genomes: The case of the missing heritability. **Nature** 456:18-21

Manjrekar J (2017) Epigenetic inheritance, prions and evolution. **J Genet** 96:445-456

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM (2009) Finding the missing heritability of complex diseases. **Nature** 461:747-753

Maynard Smith J, Szathmáry E (1995) The Major Transitions in Evolution. Freeman, Oxford

Mayr E (1961) Cause and effect in Biology - Kinds of causes, predictability, and teleology are viewed by a practicing biologist. **Science** 134:1501-1506

Mayr E (1982) The growth of biological thought: Diversity, evolution, and inheritance. Harvard University Press,

Merlin F (2010) Evolutionary Chance Mutation: A Defense of the Modern Synthesis' Consensus View. **Philos Theor Biol** 2:e103

Muller GB (2007) Evo-devo: extending the evolutionary synthesis. **Nature Rev Genet** 8:943-949

Muller GB (2017) Why an extended evolutionary synthesis is necessary. **Interface Focus** 7:11

Newby GA, Kiriakov S, Hallacli E, Kayatekin C, Tsvetkov P, Mancuso CP, Bonner M, Hesse WR, Chakrabortee S, Manogaran AL, Liebman SW, Lindquist S, Khalil AS (2017) A Genetic Tool to Track Protein Aggregates and Control Prion Inheritance. **Cell** 171:966-979

Odling-Smee FJ, Laland KN, Feldman M (2003) Niche construction: the neglected process in evolution. Princeton University Press, Princeton, NJ

Pennisi E (2008) Modernizing the modern synthesis. **Science** 321:196-197

Pigliucci M (2005) Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life. **Nature** 435:565-566

Pigliucci M, Müller GB (2010) Evolution, the extended synthesis. MIT Press, Cambridge, Massachusetts

Rendell L, Boyd R, Cownden D, Enquist M, Eriksson K, Feldman MW, Fogarty L, Ghirlanda S, Lillicrap T, Laland KN (2010) Why Copy Others? Insights from the Social Learning Strategies Tournament. **Science** 328:208-213

Richerson PJ, Boyd R (2005) Not by genes alone. University of Chicago Press, Chicago

Salt GW (1979) Comment on the use of the term emergent properties. **Am Nat** 113:145-148

Sarin S, Dukas R (2009) Social learning about egg-laying substrates in fruitflies. **Proc R Soc Lond B Biol Sci** 276:4323-4328

Shorter J, Lindquist S (2005) Prions as Adaptive Conduits of Memory and Inheritance. **Nature Rev Genet** 6:435-450

Simpson GG (1984) Tempo and mode in evolution. (eds) Tempo and Mode in Evolution. Columbia University Press

Skinner MK (2011a) Environmental epigenetic transgenerational inheritance and somatic epigenetic mitotic stability. **Epigenetics** 6:838-842

Skinner MK (2011b) Role of Epigenetics in Developmental Biology and Transgenerational Inheritance. **Birth Defects Research (Part C)** 93:51-55

Skinner MK, Anway MD (2005) Seminiferous cord formation and germ-cell programming - Epigenetic transgenerational actions of endocrine disruptors. **Ann N Y Acad Sci** 1061:18-32

Skinner MK, Guerrero-Bosagna C, Haque MM (2015) Environmentally induced epigenetic transgenerational inheritance of sperm epimutations promote genetic mutations. **Epigenetics** 10:762-771

Skinner MK, Manikkam M, Guerrero-Bosagna C (2010) Epigenetic transgenerational actions of environmental factors in disease etiology. **Trends Endocrinol Metabol** 21:214-222

Skinner MK, Manikkam M, Guerrero-Bosagna C (2011) Epigenetic transgenerational actions of endocrine disruptors. **Reprod Toxicol** 31:337-343

Slagsvold T, Wiebe KL (2011) Social learning in birds and its role in shaping a foraging niche. **Phil Trans R Soc B** 366:969-977

Sommer RJ (2004) Beyond Darwin–towards an inclusive evolutionary synthesis. **Development** 131:5769–5770

Szathmàry E, Maynard Smith J (1995) The major evolutionary transitions. **Nature** 374:227-232

Tinbergen N (1963) On aims and methods of ethology. **Z Tierpsychol** 20:410-433

Uller T, Helanterä H (2019) Niche construction and conceptual change in evolutionary biology. **British J Phil Sci** 70:351-375

van de Waal E, Borgeaud C, Whiten A (2013) Potent Social Learning and Conformity Shape a Wild Primate's Foraging Decisions. **Science** 340:483-485

Varela SAM, Matos M, Schlupp I (2018) The role of mate-choice copying in speciation and hybridization. **Biol Rev** 93:1304-1322

Vastenhouw NL, Brunschwig K, Okihara KL, Muller F, Tijsterman M, Plasterk RHA (2006) Long-term gene silencing by RNAi. **Nature** 442:882-882

Wagner RH, Danchin É, Boulinier T, Helfenstein F (2000) Colonies as byproducts of commodity selection. **Behav Ecol** 11:572-573

Wang Y, Liu H, Sun Z (2017) Lamarck rises from his grave: parental environment-induced epigenetic inheritance in model organisms and humans. **Biol Rev** 92:2084-2111

West-Eberhard MJ (2003) Developmental plasticity and evolution. Oxford University Press, Oxford

Whiten A (2005) The second inheritance system of chimpanzees and humans. **Nature** 437:52-55

Whiten A (2018) Culture and conformity shape fruitfly mating. Potent social learning sustains the inheritance of mating preferences over generations. **Science** 362:998-999

Whiten A (2021) The burgeoning reach of animal culture. **Science** 372:6537

Whiten A, Goodall J, McGrew WC, Nishida T, Reynoldsk V, Sugiyama Y, Tutin CEG, Wrangham RW, Boesch C (1999) Culture in chimpanzees. **Nature** 399:682-685

Williams GC (1966 [2019]) Adaptation and natural selection: A Critique of Some Current Evolutionary Thought. Princeton university press,

Wright S (1930) The genetical theory of natural selection: a review. **J Heredity** 21:349-356

Wright S (1931) Evolution in Mendelian populations. **Genetics** 16:97

Xia JF, Han L, Zhao ZM (2012) Investigating the relationship of DNA methylation with mutation rate and allele frequency in the human genome. **BMC Genomics** 13:S7-S15

Chapter 19

Darwinian/hennigian systematics and Evo-Devo: the missed Rendez-vous

Guillaume Lecointre[i]

Abstract

EvoDevo called homeotic genes “architect genes” because they “control” “body plans”. Using such ordering causal factors, EvoDevo stayed outside Evolution because in biology, order is not causal, it is a consequence that we need to explain. Natural selection is the concept explaining the rise of apparent order, or regularities. Genes do not control anything, they just impulse. The two pillars of evolutionary theory - descent with modification and natural selection- do occur within the organism itself and the name for it is ontophylogenesis. It should have been EvoDevo. To study ontophylogenesis and developmental heterochronies among species, we propose a hierarchical analysis of their ontogenetic time. With a parsimony analysis of a matrix where “operational taxonomic units” are species at a given ontogenetic time segment and characters are organs or structures which are coded present or absent at this time, we show that the hierarchies obtained have both very high consistency and retention index, indicating that the ontogenetic time is correctly grasped through a hierarchical graph. We consider such graphs as real phylogenies and the very core of EvoDevo. This allows to formally detect developmental heterochronies and might provide a baseline to name early life stages for any set of species (where terminologies previously used were different across species) following hennigian principles. Our phylogenetic segmentation of ontogenetic time depicts ontophylogenesis.

---

[i] Institut de Systématique, Évolution et Biodiversité, UMR 7205 CNRS-MNHN-SU-EPHE-UA, Case Postale 24, Muséum National d'Histoire Naturelle, 57 rue Cuvier, 75005 PARIS, France. Email : Guillaume.lecointre@mnhn.fr

19.1What is evo-devo for?

EvoDevo aims at understanding morphological complexity of organisms through comparing across species structures and functions of genes involved in development. Interest is not only devoted to model species, but also to "minor phyla”. Genetic determinants and their effects are mapped onto phylogenies in order to densify the taxonomic sampling of organisms and to better understand the evolutionary steps leading to extant morphologies and anatomies. Müller (2007) correctly summarized the historical context of emergence of EvoDevo:

> Evolutionary developmental biology (EvoDevo) emerged as a distinct field of research in the early 1980s to address the profound neglect of development in the standard modern synthesis framework of evolutionary theory, a deficiency that had caused difficulties in explaining the origins of organismal form in mechanistic terms.

The point we address here is that, looking at today’s EvoDevo papers, paradoxically, EvoDevo seems to partly stay outside evolution. To be more precise, EvoDevo continues to maintain a number of platonistic concepts that are not compatible with today’s evolutionary theory and could be replaced by concepts already available in the fields of phylogenetic systematics. But changes are already ongoing. *Natural selection* and *Descent with modification*, the two pillars of the Darwinian approach of life, are entering the soma (not restricted to the adult soma, but also to the entire developing soma, avoiding “adultocentrism”, Minelli 2014: 233). Today it is possible to construct the phylogeny of cells of a single developing organism (Briggs et al. 2018; Farrell et al. 2018; Chapman et al. 2020), or to perform a phylogenetic analysis of metastases from a single patient (Zhao et al. 2016). Natural selection within the body, anticipated by Roux (1881, 2013), but occulted during the past century, is now coming back, allowing to understand cancer dynamics, ageing, neurogenesis, etc. Ontogenesis and phylogenesis are no more two distinct processes: natural selection and descent with modification explain both the developing individual and its stability, as well as the regularity of individuals from which we name species. Ontophylogenesis (Kupiec 2012) should have been the ultimate goal of EvoDevo, but was born outside it.

EvoDevo incorporated into its research program idealistic views where order (things are at the "right place") and regularity (the order to come is already written) are the cause of biological phenomena. Genetic control, genetic program, architect genes, phylotypic stages, body plans do ensure order. These concepts all use the notion of instruction: the order (biological organization, regularities to come) is already written somewhere and development is the unfolding of a program. Nijhout (1990) diagnosed:

> The concepts that genes control development and morphology, that genomes contain developmental information, and that development follows a genetic program pervade modern thinking in molecular, developmental, and evolutionary biology. The genome is assumed to encode higher levels of organization. Genes and their products are seen as the causative agents of differentiation, and controlled gene expression is seen as the driving force of progressive change in development. The crucial regulatory role attributed to genes is emphasized by the widespread acceptance of the notion that a substantial number of genes are specifically concerned with the orderly progression of events during development. As a consequence, it is assumed that an understanding of the mechanisms of gene regulation and of the detailed structure of the genome are not only fundamental to an understanding of development but virtually sufficient for this understanding.

These ideas described by Nijhout have their deep roots into idealistic morphology of the German nineteenth century and preformationist thinking of the seventeenth century: the phenotypic order is sufficiently explained by a microscopic order where the homunculus is replaced by the "genetic program". The fact is, such ideas are not compatible with evolutionary biology. In modern biology, order does not explain anything, it is what we need to explain from disorder and change at the chemical scale. If the explanation of the regularity of forms between the baby cats and their cat mother is “because there is a cat genetic program”, we have explained nothing at all. Vignaux (1977: 15) already pointed out the circularity of Lwoff’s claim (1969):

> La seule source d’ordre biologique est l’ordre biologique.

Development is not the unfolding of a program (Moczek 2014: 224; Kampourakis 2017: 172-173), it is a construction (Laland et al. 2015).

Charles Darwin (1859) revolutionized biology precisely because, with the principle of natural selection, he explained apparent order (regularity of forms in a same species and fit between forms and functions) from disorder (random variation), ignorantly importing Maupertuis’s (1751: 15) intuitions into science. During the twentieth century, biochemistry, molecular biology (molecular genetics included), molecular physiology, medical research, and EvoDevo were non-darwinian, in that sense. These disciplines tried to explain biological order from biological order, ignoring that natural selection was a concept already available to explain an apparent order (or regularity) at a given scale of space and (short) time from disorder at a lower scale. Unfortunately, the Nobel prices Lwoff, Monod and Jacob tried to explain macroscopic order as a consequence of a *sufficient microscopic order* (instructions of a program). Mayr (1961) did the same by introducing the “program” and the associated concept of teleonomy to eradicate suspicion of teleology in Biology, in order to explain the apparent purposefulness of organisms and their characteristics. In a less sophisticated manner, Jacob (1970: 17) did not mention teleonomy but confessed:

> Longtemps le biologiste s’est trouvé devant la téléologie comme auprès d’une femme dont il ne peut se passer, mais en compagnie de qui il ne veut pas être vu en public. A cette liaison cachée, le concept de programme donne maintenant un statut légal.

Anywhere Biology makes a causative use of order (e.g. by using a genetic upward causation), biology thinks outside its own theory. Put in another way, evolution and above all, evolutionary thinking, must fully enter the soma and its development. EvoDevo is just starting this mutation four decades after its birth.

19.2 Genetic control

EvoDevo continues to use the verb “to control” to specify the type of influence that genes have. The genetic upward causation (Kampouralis 2017: 173) is still the main type of genetic reductionism. The most common verbs associated to genes in today’s EvoDevo papers are “to control” and “to regulate”. These instructionist expressions have probably been incorrectly reinforced by experimental knockouts of genes: when the determinant is suppressed, so is the effect. However, necessity and control are not synonymous: such effects of knockouts could have been perfectly understood under a moderate regime of causation. If one replaces “control” by “impulse”, a completely different view of development is possible, while maintaining correct interpretation of gene knockouts: when one suppresses the impulse, the subsequent cascade of events is also suppressed. In parallel, with the use of the verb “impulse”, development can be considered as a construction (Laland et al. 2015) where genes are partners, neither controllers nor managers (Nijhout 1990; Kupiec and Sonigo 2000; Moczek 2012, 2014: 223). Kampourakis (2017: 171) correctly entitles an entire chapter: “*Genes are implicated in the Development of Characters*”, which is a correct way to describe the role of genes.

Yet the need for a revision of the role we should give to genes to understand development have been pointed out for long ago. In a clear, brilliant and constructive paper, as early as 1990 Nijhout warned us that the role previously given to genes was inappropriate and counter-productive. He clearly expressed the notion that genes are not controllers, but just partners:

> When we trace the causal pathway of a developmental event, we may often (but not necessarily always) encounter a gene whose product is required for that event, and without which that event would not take place. But the causal pathway does not end

> there. The expression of the gene or the activity of its product must itself be controlled by a specific stimulus, perhaps an ionic or organic inducing molecule, or through the product of a regulatory gene. Regulatory genes, in turn, owe their timely activity to stimuli external to themselves, and so forth. The causal pathway is endless and involves not only genetic, but manifold structural, chemical and physicochemical events, a defect in any of which can derail the normal process. (…) Thus, genes do not provide instructions for development, but they aid in supplying the material basis for development. (…) Furthermore, the causal pathway is seldom if ever linear, but contains loops and complex reticulations. Thus, when we speculate about the control of a developmental process, it is misleading to assign a controlling role to a particular gene.

These reticulations, coupled with the fact that gene effects are intricated, partly explains apparent stability and repeatability of a development, along with natural selection among cells (see below). Last but not the least, Nijhout (1990) gave the most precise view of the role we should give to genes:

> The most generally useful hypothesis about the function of genes is the following: Genes are passive sources of materials upon which a cell can draw, and are part of an evolved mechanism that allows organisms, their tissues and their cells to be independent of their environment by providing the means of synthesizing, importing, or structuring the substances (not just gene products, but all substances) required for metabolism, growth and differentiation. The function of regulatory genes is ultimately no different from that of structural genes, in that they simply provide efficient ways of ensuring that the required materials are supplied at the right time and place.

19.3 Genetic program

The genetic program is a metaphor imported into Biology from cybernetics and informatics by Mayr (1961), and by Monod and Jacob (1961) (Peluffo 2015). The birth of the notion is clearly not embedded into the evolutionary theory. The metaphor accounts for the process by which regularity of forms occurs *as if* there was a program to unfold them. As a preformationist metaphor, it is no way compatible with evolutionary biology and the notion is now so amended that it became useless (Nijhout 1990; Moczek 2012; Heams 2004, 2013a). Again, as early as 1990, Nijhout provided the best and simplest arguments that the "program" is an inappropriate metaphor:

> Does the fact that in most cases gene expression in development is sequential constitute a program? Two conditions must be met for this to be true. First, the sequence from gene to process must be causal, that is, the gene or its product must be necessary and sufficient for the occurrence of the process, and not be itself provoked by the process. Without such a stipulation the relation becomes trivial; e.g. a bouncing ball consists of a sequence of causal reactions, but this does not mean that the ball is programmed to bounce, nor is it useful in an analysis of the physics of bouncing to suppose that such a program might exist. (…) For instance, actin and tubulin genes are necessary for morphogenetic movement to occur. A deficiency in these genes (or in any gene or process that affects the synthesis or localization of actin and tubulin) would prevent or severely distort much of morphogenesis. These genes do not, however, 'cause' or 'control' morphogenesis; they enable it to take place. (…) Second, a program must somehow contain information about the temporal sequence of events. This criterion is never met. Development is a series of elaborate temporal and spatial interactions that are context dependent. The sequence of gene activation we see in development is an emergent property of this interaction (again, the bouncing ball analogy). The genes whose products are necessary during development are activated by stimuli that arise from the cellular and chemical processes of development. Thus the network or pattern of gene activation does not constitute a program, it is both the consequence of, and contributor to, development.

Several retired geneticists of the “program generation” claim that attacks against the program is actually an attack against a caricature of it, that the metaphor has always been open and subtle. Indeed, right from the beginning this is what Mayr wrote in 1961:

> It is characteristic of these genetic codes that the programming is only in part rigid. Such phenomena as learning, memory, nongenetic structural modification, and regeneration show how "open" these programs are (Mayr, 1961).

However, some sociological facts have to be taken into account. If sophisticated geneticists could take the “genetic program” as a loose metaphor, it is not the case for researchers of other fields. Mayr (1984: 126) himself wrote:

> all of the directions, controls and constraints of the developmental machinery are laid down in the blueprint of the DNA genotype as instructions or potentialities.

According to Gehring (1984), the genome contains a developmental program that is not so “open”:

> It has become increasingly clear that the developmental program resides in the genome, and that in most cases the environment provides only general stimuli and relatively little specific information.

Here the metaphor is not an analogy anymore but describes the mechanism itself. Moczek (2012) also pointed out that Biology textbooks do not take the notion of genetic program just as a metaphor. Moreover, even if the term “genetic program” seems to have been taken as an analogy, the meant instructions dictated by the program clearly were not. For instance, let's read Jacob (1970: 10):

> Ce qui est transmis de génération en génération ce sont les « instructions » spécifiant les structures moléculaires. Ce sont les plans d'architecture du futur organisme. Ce sont aussi les moyens de mettre ces plans à exécution et de coordonner les activités du système. Chaque œuf contient donc, dans les chromosomes reçus de ses parents, tout son propre avenir, les étapes de son développement, la forme et les propriétés de l'être qui en émergera. L'organisme devient ainsi la réalisation d'un programme prescrit par l'hérédité.

In their textbook, Watson et al. (1987) claimed:

> We know that the instructions for how the egg develops are written in the linear sequence of bases along the DNA of the germ cells.

Instructions are not metaphoric anymore. Things do not occur *as if* there were instructions: there *are* instructions. Such instructions led Biology to be gene-centered for half a century, where genes were not seen as partners, but controllers. Natural selection did not need to enter the soma, because the source of stability, regularity, order was already coded into a *written* program. Then there were two theoretical streams in biology: the unfolding of a program explained the rise of individual -ontogeny-, natural selection explained similarity among individuals (by pruning generations, eliminating extreme variants), from which we create species. The origin of individuals was thought through a causative order (program's instructions), the origin of species was thought from pruning a causative disorder (Kupiec, 2012). Evo-Devo was born partly because of this schizoid state of the art. There was a need of a link between two biological theories. This state of the art is the reason why there is still, as Müller (2007) puts it, "Evo-Devo questions" (like "how did the developmental repertoire evolve?") and "Devo-Evo questions" (like "How does development contribute to phenotypic novelty?” or “How does development influence phenotypic variation?”). Because EvoDevo did not push its research program until ontophylogenesis (Kupiec, 2012), proposing a single theoretical foundation for the developing organism and for the origin of species, these questions stayed and remain separated. In other words, if development had been interpreted as an evolutionary process right from the beginning, there would not have been any EvoDevo.

Unfortunately, Mayr, Lwoff, Jacob and Monod did not think as evolutionists, but as informaticians. In a letter that Mayr wrote to Jacob in 1991, and published by Peluffo (2015), Mayr implicitly recognizes (1) the vitalist structure of the “program” thinking and (2) the borrowing from computer sciences:

> I have just finished a small piece on vitalism; you are quite right, the vitalists were the ones who established the autonomy of teleology. And if one replaces the words vis, vita, . . . entéléchie, etc. ... in their discussion with the words “genetic program,” most of what they say makes perfectly good sense. I wish we had a better history of the concept of the genetic program. The word program, of course, comes from the computer language, but surely there was something that preceded it.

According to Kupiec (2014: 155), such an instructionist way of thinking has its roots in a physician’s mind. In “*What is life*”, Shrödinger (1944) wrote:

> Let me use the word “pattern”of an organism in the sense in which the biologist calls it “the four-dimensional pattern,”meaning not only the structure and functioning of that organism in the adult, or in any other particular stage, but the whole of its ontogenetic development from the fertilized egg to cell to the stage of maturity, when the organism begins to reproduce itself. Now, this whole four-dimensional pattern is known to be determined by the structure of that one cell, the fertilized egg. ...It is these chromosomes, or probably only an axial skeleton fibre of what we actually see under the microscope as the chromosome, that contain in

> some kind of code-script the entire pattern of the individual's future development and of its functioning in the mature state. Every complete set of chromosomes contains the full code. ...But the term code-script is, of course, too narrow. The chromosome structures are at the same time instrumental in bringing about the development they foreshadow. They are law-code and executive power—or, to use another simile, they are architect's plan and builder 's craft —in one.

Note that at these times the uses of the term "program" and the term "code" were interchangeable (Peluffo 2015). Architect's plan is right from the beginning into chromosomes, therefore the egg. This is preformationism.

19.4Architect genes

If there is an architect plan, there are architect genes. Hox genes are often called "architect genes", even sometimes "choregraphers" (e.g. see Philippidou and Dasen 2013). No need to develop further, except by citing Moczek (2012):

> (…) gene- and genome-centric views of development and developmental evolution are unrealistic and unproductive.

19.5Phylotypic stage

Animal phyla are taxonomic conventions. Consistent with idealistic thought by which concepts and ideas have ontological priority over real individuals, evodevo tried to justify animal phyla taken *a priori* with the hourglass model of development which was supposed to account for the unique "body plan" of each phylum. As correctly noticed by Hejnol and Dunn (2016):

> There have been many attempts in EvoDevo to provide a biological definition for what a phylum is, but all have been unsuccessful and circular. For example, a phylum is often defined as a clade with a “unique body plan” and a “unique body plan” is defined as features that are specific to a phylum.

Sure, phyla are remnants of systematics of the times of the evolutionary synthesis, suffering from some degree of taxonomic realism (Lecointre 2015a: 196-197). Researchers of those times looked for interrelationships among groups (phyla, classes, etc.) that were conceived *a priori*, allowing old linnean classes (possibly paraphyletic, like fishes) to survive (“groups, then relationships” research program). Hennig’s phylogenetic systematics just did the contrary: its research program was to look at the interrelationships among members of the content first, and define monophyletic groups (= clades) *a posteriori* (“relationships, then groups” research program). For modern hennigian systematics, phyla, if any, are just clades like others, and as such should be defined with synapomorphies or else must be dismembered (Lecointre 2015a). With the phylotypic stage, EvoDevo maps onto modern phylogenies a pre-phylogenetic way of thinking, sometimes leading to circular reasonings even in the most visible journals, like in Levin et al. (2016), which is a symptom of the degree of acceptance of these epistemological bias among the EvoDevo community. The phylotypic stage is an attempt to force the non-evolutionary concept of “body plan” to be *a posteriori* justified by phylogenies with which they are incompatible.

19.6Body plans

A body plan is a crystallization of an assemblage of characters and their mutual connections into an *invariant*. EvoDevo continues to use the concept, as each year papers using it do come out (for instance, Levin et al. 2016; Woltering et al. 2018). Invariants can’t be a reliable source of thinking in Biology, because Biology is the science of variation, even within a developing organism. Let’s recall that fact with Nijhout (1990):

> (…) at some level of microanatomy there is enormous variability in the pattern of almost everything in development; the precise pattern of cellular events at gastrulation or during osteogenesis is very variable, as is the cell-level anatomy of almost all structures from leaves to limbs. Yet at higher levels of organization the variability in these systems is greatly diminished.

If one wants to point out a regularity of forms across individuals, the subtle concept of primary homology is already available (de Pinna 1991; Lecointre 2015a). We don't need invariants in Biology (Lecointre 2017) because Biology has no law (Gayon 2003), just principles. Body plans were used as a mnemotechnic way to learn one's zoological lesson, but shortly and often they pass from pedagogy to research, with the idealistic ontological priority given to mental schemes (Lecointre 2015a: 170). Body plans do not exist outside the need for a student to memorize its lesson. Each organ can be lost during the course of evolution. The vertebrate body plan includes the possession of eyes, vertebrae, appendages, tail, among other features. The olm (*Proteus*) and the Mexican blind cavefish (*Astyanax mexicanum*) don't have eyes. Hagfishes don't have vertebrae. Snakes and several lineages of lizards don't have any appendages. The sunfish, caecilians and frogs don't have any tail. All these organisms contradict the vertebrate "body plan". Many researchers would save the concept by saying that these organisms *have lost* these anatomical characters. By justifying the body plan that way, they pass from descriptive criteria used for identification and learning to scientific inference. Just because evoking the loss of anatomical characters in past genealogical lineages is clearly speaking about evolution. They pass from a theory-free mnemotechnic notion of body plan to its intrusion into the theory of evolution. The loss of organs allows the plan to softly break in theory of evolution with which it is incompatible. During evolution, everything can indeed be lost or transformed, contradicting the "plan" itself. We can't justify a notion (the body plan) by using a theory (evolution) with which it is incompatible (Lecointre 2015a: 173). The body plan is not the appropriate concept. Phylogenetics invite us to see organisms as mosaics. At each node of the tree of life, there are no body plans at all, there are synapomorphies, i.e. formal justifications under the form of exclusive derived attributes. Hypothetical ancestors are not body plans, but they are incomplete, labile jigsaw puzzles.

19.7Order and regularity: causes, or consequences?

During the last third of the XXth century the genetic program accounted for the regularity and repeatability of developments. Other processes such as self-organization and the stabilizing effects of intricated networks of genetic influences have been evoked to account for their stability (Guo and Amir 2021) and repeatability (i.e. robustness). In parallel, in biological lessons, natural selection was taught as a factor of change. Natural selection was supposed to explain how a species changes over time. To account for a repeatable developing organism, researchers did not need a factor of change, but factors of stability and robust repeatability. The genetic program played that role. Then, during half a century we have taught that the regularity of species was ensured by the genetic program, and the change of species by natural selection. This view is a consequence of species realism, maintained, among others, by Mayr. Mayr gave priority to understand what a species is, here and now in synchrony, privileging processes over patterns, leading to its "biological concept" of species. Species being the given, so one has to explain how the given species changes. Other authors like Simpson gave priority to understand species in diachrony, privileging patterns over processes. The first approach tends to species realism (the given is species), the second favors a perception of species as a linguistic convention. In this second approach, the given being the varying individuals, then one has to explain how a species does not change: this was Darwin's approach. Today, the theoretical phylogenetic definition of species is a set of individuals being members of a same genealogy as long as this genealogy is not splited. If there is a split, whatever the reason, another species name must be given to subsequent daughter branches. Empirically speaking, a species here and now is just a hypothesis made by taxonomists. It is the hypothesis that all known members are parts of a same isolated genealogical lineage. This hypothesis is supported by several criteria, the most common being similarity and interbreeding, and is made to facilitate language and communication. Modern phylogenetic concept of species is nominalist: what do exist are individuals.

Going back to Darwin (1859), species are conventions aimed to name a certain degree of similarity among individuals. What is given is not species itself, but individuals (Table 1). This is the reason why he could pay attention to variations among them (contrary to Linnaeus who explicitly neglected variation). Darwin offered a nominalist explanation of the Origin of species by asking the question: given the variation among individuals that do interbreed, what is the cause of similarity among individuals? Natural selection, on the short term, explains similarity: each generation is pruned, extreme variants being eliminated. From the resulting similarity we do create species for the needs of our language. Thus, for Darwin, first of all natural selection acts as a stabilizing factor. We just have to consider the profound meaning of the subtitle of his main book: "*The origin of species by means of natural selection or the preservation of favoured races in the struggle for life*". The words evolution, transformation or transmutation are absent. The word "preservation" is used to specify the maintain of something. Obviously, on the long term if the environment changes, the mean form of the species will change. But first of all, natural selection explains apparent stability, similarity, i.e. regularity across individuals. In other words, apparent order is the short-term consequence of natural selection.

This was neither fully understood nor taught during the last third of the last century because the genetic program replaced natural selection within the soma to account for its stability.

***Table 1**: Effects of Mayr's species realism, after Lecointre (2015b, c).*

Century:	**What explains:**	**What is to be explained:**	**The given:**
XVIIIe Linnaeus	God	Regularity	Species
XIXe Darwin	Natural Selection	Regularity and change	Individuals
XXe Mayr	Natural Selection	Change	Species
XXIe	Natural Selection	Regularity and change	Individuals

19.8Natural selection within soma

The cause of regularity and “fine tuning” of somatic functioning was thought as the result of a program. It is time to replace the program by natural selection among cells, without excluding the stabilizing effects of intricated consequences of many genetic impulses (“genetic networks”) and, at certain molecular levels, self-organization. Roux (1881) imported natural selection within the organism. Darwin read the book a year before his death and declared in a letter to G.J. Romanes (Heams 2013b):

> As far as I can imperfectly judge, it is the most important book on evolution which has appeared for some time

A century later, natural selection was locally imported in some somatic processes described by neurosciences (e.g. Edelman’s “neural darwinism” in 1987, and his opposition to instructionist approaches both to immunology and to neurociences). Immunology has also been one of the first biological fields to take its distances with instructionist schemes of explanation, as early as 1966 (e.g. Brenner and C. Milstein used random somatic hypermutation of immunoglobulin genes to explain immunoglobulin diversity, G. Edelman and J. Gally used random somatic gene recombinations as a source of immunoglobulin diversity, MacFarlane Burnett’s “clonal selection” being different than what we call here natural selection). Cancerology did it soon after (P. Nowell in 1976; Sonnenschein and Soto 1999, Soto and Sonnenschein 2005). Soto et al. (2008) summarized the role physicalism and downward causation in developmental and cancer biology of the past century. Above all, at the very end of the last century two major theoretical advances took place. First, variation (and not instruction) at the source of any phenomenon occurring within and among cells -already anticipated by Edelman- was generalized when gene expression itself began to be understood as stochastic, culminating with the remarkable study of Elowitz et al. in 2002 (Raj 2008; Heams 2013c). Second, Kupiec (1997) based his views on fundamental stochasticity of gene expression to propose a Darwinian theory of cell differentiation (Kupiec 2014). Today random variation among cells appears to have a better explanatory power than instructionist models to understand the development and the functioning of an organism (Kupiec et al. 2013). As a result, the study of cancer and ageing now fully adopt models involving natural selection

(Nelson and Masel 2017) and metastasis is considered as an evolutionary process (Turajlic and Swanton 2016). Grajzel et al. (2020) perfectly summarized the present state of the art by the formula:

> Cancer is a genetic disease fueled by somatic evolution.

Somatic evolution is not only for cancerous cells, but for all cells. Among them, we find somatic variation (indeed at very high degree among tumoral cells), transmission (through mitosis) and constrains (nutrients, space…). These are the three fundamental conditions to obtain natural selection of some cellular lineages over others. In the past decade cancer and evolution made front pages of the most visible journals (Willyard 2016) and therapies based on evolutionary reasonings are developed and appear to be successful (Enriquez-Navas et al. 2016; Degregory and Gatenby 2019ab; Thomas 2019).

19.9Descent with modification (then Phylogeny) within soma

Now it is possible to trace a phylogenetic tree of different metastases of a single patient (see for instance Zhao et al. 2016). The comparison of transcriptomes of single cells allowed to reconstruct the phylogeny of tissues of a developing zebrafish (Farrell et al. 2018; Wagner et al. 2018), or the phylogeny of tissues of the frog *Xenopus* (Briggs et al. 2018). Interestingly, the tissues that we had learnt to be homogeneous in origin (endoderm, mesoderm, ectoderm) actually are not. In the frog, ectoderm is paraphyletic and mesoderm is monophyletic, while in the zebrafish ectoderm is monophyletic and mesoderm is paraphyletic. Surprisingly, there are several ways to develop a vertebrate: in *Xenopus* ectoderm is diphyletic and the mesoderm cell lineage well separated from the rest, while in the zebrafish ectoderm lineage is well apart and mesoderm gives birth to endoderm. This is a strong experimental argument against the vertebrate body plan, and the notion of "body plan" in general; the true phylogenetic reasoning invites us to view patterns and processes as evolutionary mosaics. More recently, the lineage tracing of human development was obtained through the phylogenetic analysis of somatic mutations (Chapman et al., 2020), allowing to discover the hypoblastic origin of extra-embryonic mesoderm and primitive blood. Schmid-Siegert et al. (2017) could reconstruct the phylogeny of somatic mutations in an oak. Actually, the view of a phylogeny reconstructed from parts of a single organism had been initiated for long ago by Fitch (1970). By defining orthologous genes, paralogous genes and metalogous genes among different copies of a multigenic family (e.g. globin genes), he already conceived that a phylogeny of different elements of an individual could be constructed.

19.10 What EvoDevo should study: ontophylogenesis

If natural selection and descent with modification -the two pillars of Darwinian evolution- are now entered into the soma and explanative of the developing body, then development is understood as an evolutionary process. EvoDevo becomes EvoEvo, in a way. Moczek (2014) formulated this idea as an epistemological program: "*development should be nested within a theory of developmental evolution*". Kupiec (2009, 2012) already developed the idea that ontogenesis and phylogenesis are two facets of a same general process of life deployment and diversification called ontophylogenesis (Kupiec 2009, 2012). As already mentioned above, there is no process of cell deployment in the individual development ontologically separated from the process of deployment of a species. Ontogeny and phylogeny are a single process of diversifying lineages of entities that are submitted to natural selection. In absence of any platonicist invariants like "the genetic program" or "the body plan", and in absence of reification of species, phyla or individuals, EvoDevo would already have achieved Moczek's program of "*building a theory of developmental evolution*", which is ontophylogenesis (Table 2).

***Table 2**: (after Lecointre et al. 2020)*

Reified entity	Components	What is to be explained	What explains
Individuals	Cells	Ontogenesis	Genetic program
Species	Individuals	Phylogenesis	Descent with modification
None	Cells and individuals	Ontophylogenesis	Natural selection and descent with modification

19.11 A possible consequence of a true evo-devo research program: hierarchical (phylogenetic) ordering of ontogenetic time

Under the research program called Phylogenetic Systematics, Hennig (1950, 1966) provided the reasoning to construct a phylogeny as a scheme of arguments for a given set of sister-group relationships, and to produce classifications consistent with it (Lecointre 2015a). The scheme is under the form of a nested set of arguments called synapomorphies (then providing a nested set of taxa, on the systematic side). Using an evolutionary vocabulary, they are called “secondary homologies” (De Pinna, 1991), or to make it short, homologues. EvoDevo and phylogenetic systematics never really met, as if there was a sort of *missed Rendez-Vous*. This is due to the organization of disciplines during the last century. EvoDevo mostly used trees called “phylogenetic” from molecular systematics, which were built from aligned sequence data most often analyzed through phenetic methods (which are not Hennig’s phylogenetic methods, Lecointre 1997; 2015a: 189) or later through probabilistic methods. EvoDevo never constructed a nested hierarchy of secondary homologues directly from primary homologues themselves, which is the very core investigation of phylogenetic systematics. This is what Lecointre et al. (2019) have done.

Moczek (2014) advocated for a research program (that should be the one of EvoDevo) he summarized as follows:

> Step 1 would accumulate the knowledge base of development, focussed on identifying and linking developmental products and processes. Step 2 would organize this information using a three-layered approach, focussed on the development of homologues (layer 1), a nested hierarchy of homologues (layer 2), and a description of patterns and causes of variation within homologues (layer 3). The strategy outlined here would allow a conceptualization of development that (i) is biologically realistic, (ii) can be refined alongside a growing knowledge base, (iii) can incorporate both homology (descent) and variation (modification), and (iv)

is capable of bridging to relevant conceptual frameworks in adjacent biological fields.

Let’s explore step 2 layer 2.

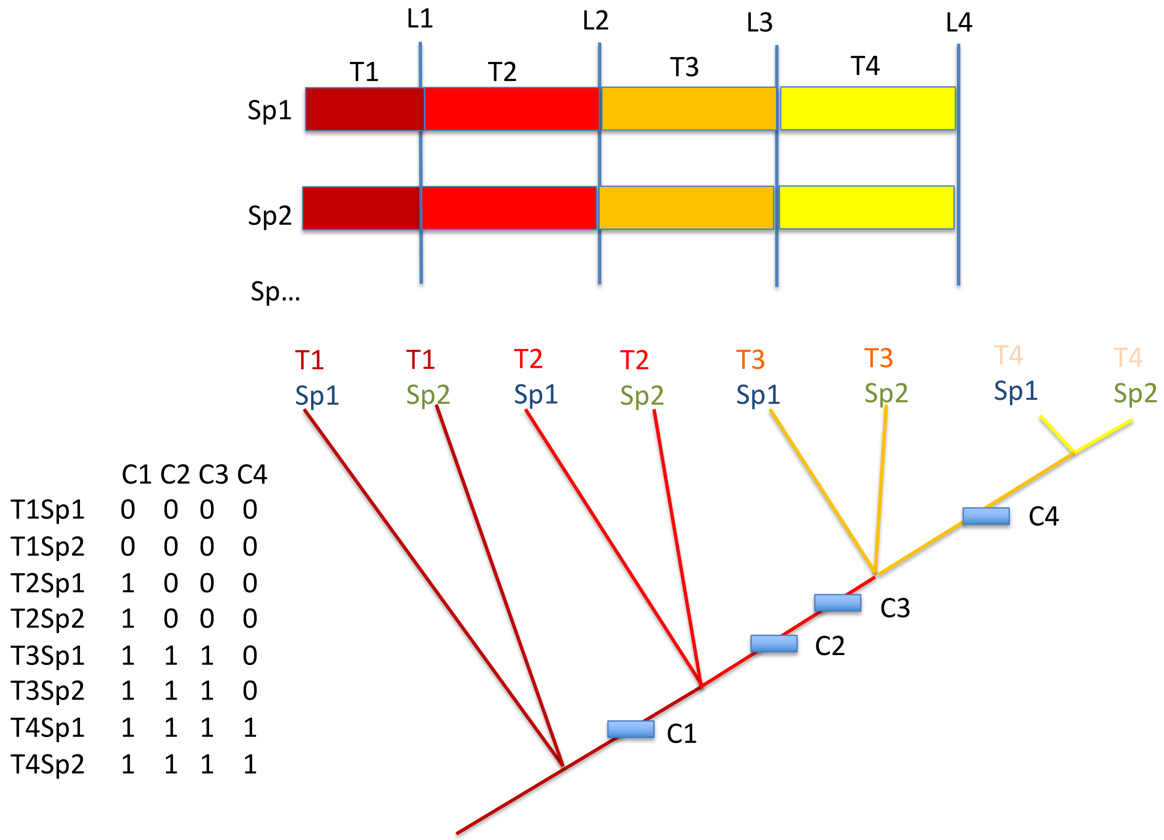
If we avoid reification of individuals and species, there is no ontological difference between a phylogeny comparing parts of a single organism and a phylogeny comparing parts of different species onto which we have made hypotheses of primary homologies (the classical phylogenetic approach). How can one compare parts of a single organism? It can be done in synchrony because there are similar parts within a same body, e.g. comparing paralogs of a same multigenic family, or comparing similar elements candidates to serial homology. Or it can be done in diachrony by comparing parts of different species at different moments of their development. Phylogeny, then, can be used to compare organs of different species at different stages. What is the benefit? This is the best method to detect developmental heterochronies. This is also the most consistent way to name developmental stages, valid of a set of different organisms. This is not a detail: what we call a “larva” is an organism at very different steps of its development according to the species we consider, even among close families of teleostean fishes. There is no consistent definition of what a larva is across the animal diversity (Haug 2020). In other words, we can’t name segments of ontogenesis with a certain degree of generality beyond the taxonomic level of species or family, a situation which severely hampers comparative anatomy and aquaculture. Let’s explain these two benefits successively.

19.11.1The method

For different species (in Lecointre et al. 2019 four teleostean freshwater species), onsets of a panel of organs are compared as well as organs presence or absence at a particular time. A developmental time frame for all species is arbitrarily segmented in five parts, defined using time landmarks: L1=0% of the developmental time, L2=25%, L3=50%, L4=75% and L5=100% (Figure 1, developmental time being normalized in degrees-Celsius-days). The 0% boundary is fecundation, the 100% boundary was arbitrarily chosen here as the onset of pectoral fin rays. Between time landmarks, there are four time-segments T1 (from 0% to 25%), T2 (from 25% to 50%), T3 (from 50% to 75%) and T4 (from 75% to 100%). The time frame 25%-50%-75% was arbitrarily chosen, but one could have chosen any other time frame. As an example, Lecointre et al. (2020) have also chosen to explore more precisely the earliest moments of development with the time frames 5%-10%-50% and 5%-10%-20%.

Then "operational taxonomic units" (OTUs) are a given species considered at a given time segment. T1Sp1 in Figure 1 is the species n°1 considered at the time segment n°1, T2Sp2 is the species n°2 considered at the time segment n°2, T3Sp2 is the species n°2 considered at the time segment n°3, etc. For each OTU, the presence (coded 1) or the absence (coded 0) of a given organ, or structure, is recorded in columns of a data matrix (Figure 1), these columns being classically called "characters" in any morphological data matrix devoted to phylogenetic analysis.

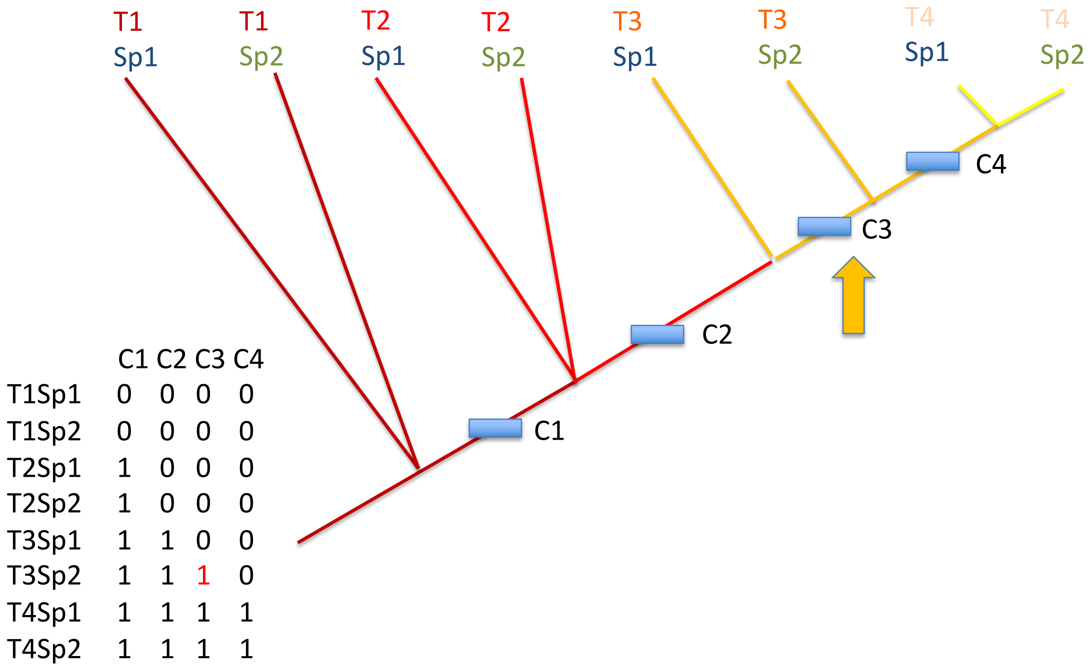
As most organs appear in a cumulative manner (e.g. the onset of skin pigmentation does not need to correspond to the loss of the heart, which itself does not need to correspond to the loss of the notochord), the nested hierarchy provided by a tree (technically speaking, a non-cyclical oriented connected graph) is suitable for ordering organ onsets. The matrix is then analyzed through a classical parsimony procedure, which maximizes the contiguity of branches having the same character states. The most parsimonious tree (in other words the most consistent hierarchy) is obtained from characters considered unweighted and unordered. As development is a cumulative process through time, the hierarchy provided by the most parsimonious (= the most consistent) tree is a time hierarchy. Therefore, it is logical to define the outgroup (the root of the tree) at the boundary 0%: outgroups will be OTUs at 0% of their development (no traits). Potential loss of organs or traits during the development will appear as character reversals, which is classical in such analyses.



	C1	C2	C3	C4
T1Sp1	0	0	0	0
T1Sp2	0	0	0	0
T2Sp1	1	0	0	0
T2Sp2	1	0	0	0
T3Sp1	1	1	1	0
T3Sp2	1	1	1	0
T4Sp1	1	1	1	1
T4Sp2	1	1	1	1

***Figure 1: General methodological framework**. Colored bars are developmental time for species 1 (Sp1) and species 2 (Sp2), and more species. L1, L2, etc. are arbitrary time landmarks measured as percentage of time (in degrees Celsius-days of development) of the full development from fecundation (0%) to the rise of lepidotrichia in pectoral fin rays (100%). Time segments T1, T2, etc. are defined between landmarks. The matrix at bottom left records presence and absences of various organs and traits as characters (columns: C1, C2, etc.) for each Operational Taxonomic Unit (line). An OTU is a given species at a given time segment. Bottom-right, the hierarchy of developmental time depicted with an oriented non cyclic connected graph (which is usually called a "tree") obtained through a parsimony analysis of the matrix. This theoretical tree shows the same relative timing of the onset of organ for the two species. After Lecointre et al. (2020).*

If all developments among species follow the same timing of events (same relative timing in organ onsets, or "synchronic" development among species), we should theoretically obtain the tree shown in Figure 1. Characters (i.e. traits or organs) are gained according to the same time hierarchy for all species. If a single heterochrony appears, for example species 1 is late at time 3 in gaining character 3 compared to all other species, one would obtain the tree shown in Figure 2. Species 1 at time 3 does not have yet the character C3 that all other species (here just species 2) already have at that time.



	C1	C2	C3	C4
T1Sp1	0	0	0	0
T1Sp2	0	0	0	0
T2Sp1	1	0	0	0
T2Sp2	1	0	0	0
T3Sp1	1	1	0	0
T3Sp2	1	1	1	0
T4Sp1	1	1	1	1
T4Sp2	1	1	1	1

Figure 2*: Same tree as Fig.1, but with an heterochronic event. The relative timing of the onset of characters for the two species is not the same, as species 1 is late in gaining character 3 compared to species 2 which already has it at time segment 3. After Lecointre et al. (2020)*

19.11.2Detection of heterochronies

In their paper, Lecointre et al. (2020) detected several heterochronic events among the four species analyzed, we report only one of them here. The parsimony analysis of the 53 characters results in two equi-parsimonious trees of length of 62 steps (strict consensus shown in Figure 3), with a consistency index of 0.85 and retention index of 0.96. Such high consistency values mean that the developmental time is a hierarchical time, in other words (1) the rise of organs is overall cumulative and (2) their timing is much similar between the four species. From Figure 3 we see that at 25% and 50% of their developmental time, *Hucho hucho* was late compared to *Thymallus thymallus*; grayling is late compared to *Tinca tinca*; tench is late compared to *Barbus*

barbus. It is interesting that this trend is modified at 75% of the developing time, where *Tinca tinca* is late (white arrow in Figure 3): at that time segment, it does not yet exhibit characters 46, 49 and 52 in contrast to all other species (46 is the onset of lepidotrichia in dorsal fins, 49 is the onset of lepidotrichia in anal fin, 52 is the onset of lepidotrichia in pelvic fins).

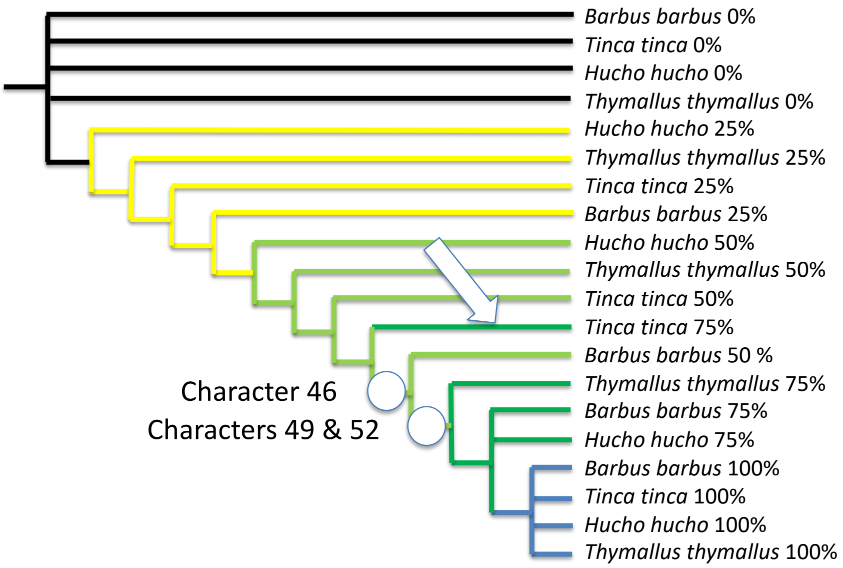
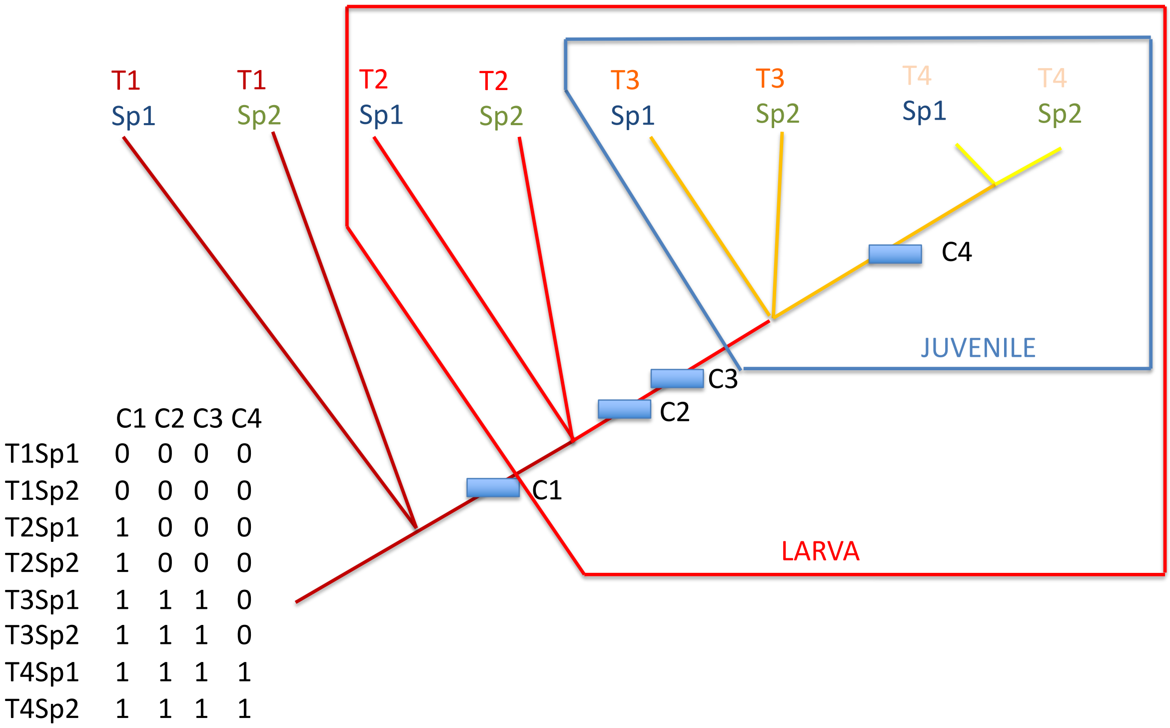


***Figure 3**: Strict consensus of two equi-parsimonious trees with the length of 62 steps, consistency index of 0.85 and retention index 0.96, obtained for the four species under the time frame 25–50–75%. Note that Hucho is late compared to Thymallus, Thymallus is late compared to Tinca, and Tinca is late compared to Barbus. The arrow shows heterochrony (see text); onsets of characters shown with circles are those detecting it. After Lecointre et al. (2020).*

19.11.3Naming developmental stages

Another outcome is a staging terminology. Even though an accurate description of early life stages is available for some teleostean species in form of embryonic and post-embryonic developmental tables, there is poor overlap between species-specific staging vocabularies beyond the taxonomic family level. What is called "embryonic period", "larval period", "metamorphosis", or "juvenile" is anatomically different across teleostean families. This problem, already pointed out 50 years ago, challenges the consistency of developmental biology, embryology, systematics, and hampers an efficient aquaculture diversification. Here the possibility to name stages with a wider validity, because one chooses the taxonomic scope of his comparison. In Figure 4, Lecointre et al. (2020) proposed a way to name stages in a hennigian way. A given node could be chosen to propose an arbitrary name. It is important to emphasize that (1) those names are valid for a wide taxonomic scope (actually the one covered by the species sampling) because based on an explicit and formal comparative method and (2) as the developmental time is a hierarchical time, stages are defined as nested sets. Indeed, to be self-consistent, a concept (a developmental stage) must contain all the entities that have the attributes by which it has been defined. For instance, it would not make sense to justify the set of mammals by the sharing of hairs and a single jaw-bone, while leaving some entities having hairs and a single jaw-bone outside mammals. Consistency of our language depends on completeness of our concepts. In traditional ways to segment time, the larva was defined by having the organ X, then the juvenile was defined by having later another organ Y (non-homologous, i.e. somewhere else in the organism). Doing so, the animal, when considered as a juvenile, was no longer considered as a larva. It is inconsistent, because the organ X by which the larva was defined is most often still there in the juvenile, and even in the adult (such as eyes, fin rays…). Developmental stages should not be segmented in an exclusive manner as series of successive sets, but segmented in an inclusive manner, i.e. should be designed as nested sets, in line with the logic of phylogenetic systematics. It is a gain in conceptual consistency, and not a loss in usefulness. Indeed, an individual of the species 2 at the time 3 (Figure 4), would be considered a juvenile. The fact that it is also still a larva is just a cumulative, ancestral property that does not need to be mentioned again. Species 1 at time 2 (Figure 4), would be considered a larva, and by saying that, one would mean that it is not yet a juvenile. Taxonomy is made to improve the self-consistency of our language. Deciding what a larva is, or what a juvenile is, are arbitrary conventions: we expect them to be self-consistent and useful.



	C1	C2	C3	C4
T1Sp1	0	0	0	0
T1Sp2	0	0	0	0
T2Sp1	1	0	0	0
T2Sp2	1	0	0	0
T3Sp1	1	1	1	0
T3Sp2	1	1	1	0
T4Sp1	1	1	1	1
T4Sp2	1	1	1	1

Figure 4: *Same theoretical tree as Fig.1, suggesting nested sets of developmental stages (see text). After Lecointre et al. (2020).*

19.12 Conclusion

Müller (2007) reviewed how EvoDevo's results "*take evolutionary theory beyond the boundaries of the Modern synthesis*", which is a way of promoting EvoDevo's scientific fecundity, for good reasons. But there is still a paradox: EvoDevo continues to publish papers full of instructionist and/or platonician metaphors. Sometimes metaphors are productive for a certain time span, and finish to become an obstacle later. As nicely summarized by Peluffo (2015):

> However, metaphors that "illuminate matters quickly and efficiently" may dim "with time and frequent usage" (Wilkins 2013) until they no longer capture the complexity of the field to which they belong.

EvoDevo should fully enter into evolutionary thinking to which, paradoxically, its contributes: contrary to Lwoff (1969), biological order does not come from biological order. Biology is neither physics nor chemistry. Biology is Biology because its explanations do not deal with *invariants* like universals and laws, but with *historical singulars* and their variations. To do so, EvoDevo should abandon the platonician metaphors of body plan, phylotypic stage, and the associated taxonomic realism, which are clearly potential sources of methodological bias (see Levin et al. 2016 and Hejnol and Dunn 2016); as well as the instructionist notions of genetic program and genetic control. It should be replaced by a true phylogenetic way of thinking (the diversity of organismal patterns deploys through time as mosaics, not plans), introduce phylogeny into the soma to depict its development, and natural selection as a stabilizing source, just as Darwin did to explain the regularity of individuals from which we create species. EvoDevo is the right place to unify Biology. There should not be two separate theories in biology, the one explaining the rise of the reified individual through the unfolding of an instructionist program, and the other explaining the rise of a reified species through natural selection. The achievement of the EvoDevo program should be ontophylogenesis, where *descent with modification* and *natural selection* enter into the somatic development from egg to death. This change is ongoing, better explaining cancers and ageing, among other phenomena occurring within this fascinating consortium of cells that a metazoan is.

References

Briggs, J.A. et al. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* 360(6392): eaar5780. https://doi.org/10.1126/science.aar5780.

Chapman, M.S. et al. (2021). Lineage tracing of human development through somatic mutations**.** *Nature* 595(7865), 85-90. doi: 10.1038/s41586-021-03548-6.

Darwin, C. (1859). *The Origin of Species* (1st edition, Penguin Classics, 1985). Penguin Books.

De Pinna, M.C.C., (1991). Concepts and tests of homology in the cladistic paradigm. *Cladistics* 7(4), 367-394.

Degregory, J. and Gatenby, R.A. (2019a). Darwin's Ideas on Evolution Drive a Radical New Approach to Cancer Drug Use. *Scientific American* 321(2), August 2019.

Degregory, J. & Gatenby, R.A. (2019b). La théorie de l'évolution, nouvelle arme contre le cancer. *Pour La Science* 505, 26-32.

Elowitz, M.B. et al. (2002). Stochastic Gene expression in a single cell. *Science* 297(5584), 1183-1186.

Enriquez-Navas, P.M. et al. (2016). Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer. *Sci. Transl. Med.* 8(327), 327ra24. https ://doi.org/10.1126/scitranslmed.aad7842.

Farrell, J.A. et al. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* 360, 979-986. https ://doi.org/10.1126/science.aar3131.

Fitch, W.M. (1970). Distinguishing homologous from analogous proteins. *Syst Zool.* 19(2), 99-113. PMID: 5449325. https ://doi.org/10.2307/2412448.

Gayon, J. (2003). « Il était une fois… ». *Sciences et Avenir* hors-série 134, 16-21.

Gehring, W.J. (1984). Homeotic Genes and the Control of Cell Determination. *In*: E.H. Davidson and R.A. Firtel (eds.), *Molecular Biology of Development* (p. 3-22). A. R. Liss, New York.

Gould, S.J. (1977). *Ontogeny and Phylogeny*. Belknap Harvard University Press.

Grajzel, D., Derényi, I. and Szöllősi, G. J. (2020). A compartment size-dependent selective threshold limits mutation accumulation in hierarchical tissues. *Proc. Natl. Acad. Sci. USA* 117, 1606–1611. https ://doi.org/10.1073/pnas.19131 04117.

Guo Y., and Amir, A. (2021). Exploring the effect of network topology, mRNA and protein dynamics on gene regulatory network stability. *Nat Commun.* 12(1), 130. doi:

10.1038/s41467-020-20472-x. Erratum in: *Nat Commun*. 12(1), 994. PMID: 33420076; PMCID: PMC7794440.

Haug, J.T. (2020). Why the term “larva” is ambiguous, or what makes a larva? *Acta. Zool.* 101, 167–188. https ://doi.org/10.1111/azo.12283.

Heams T. (2004). Biologie moléculaire : affronter la crise de la cinquantaine. *In:* J. Dubessy, G. Lecointre and M. Silberstein (eds), *Les matérialismes (et leurs détracteurs)* (p. 237-261). Syllepse, Paris.

Heams, T. (2013a). Existe-t-il un programme génétique ? *In:* J.J. Kupiec (ed.), *La vie, et alors ?* (p. 131-146). Belin, Paris.

Heams, T. (2013b). La lutte des parties dans l’organisme ou l’impasse visionnaire. *In*: T. Heams (ed.), L. Cohort, S. Danizet-Bechet, A.L. Pasco-Saligny and C. Thébault (trad.), W. Roux, *La lutte des parties dans l’organisme* (p. 9-25). Éditions Matériologiques, Paris.

Heams, T. (2013c). Expression stochastique des gènes et différenciation cellulaire. *In*: J.J. Kupiec, O. Gandrillon, M. Morange and S. Silberstein (eds), *Le hasard au cœur de la cellule* (Second edition, p. 31-62). Éditions Matériologiques, Paris.

Hejnol, A. and Dunn, C.W. (2016). Animal Evolution: Are Phyla Real? *Curr Biol.* 26(10): R424-6. doi: 10.1016/j.cub.2016.03.058. PMID: 27218852.

Hennig, W. (1950). *Grundzüge einer Theorie der phylogenetischen Systematik*. Deutscher Zentralverlag.

Hennig W. (1966). *Phylogenetic Systematics*. University of Illinois Press, Urbana.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3, 318-356, DOI 10.1016/S0022-2836(61)80072-7

Jacob, F. (1970). *La logique du Vivant*. TEL Gallimard, Paris.

Kupiec, J.J. (1997). A Darwinian theory for the origin of cellular differentiation. *Molecular and General Genetics* 255(2), 201-208. https://doi.org/10.1007/s004380050490

Kupiec, J.J. (2009). *The origin of individuals*. World Scientific, Singapore.

Kupiec, J.J. (2012). *L'ontophylogenèse. Évolution des espèces et développement de l'individu.* Collection Sciences en questions, Quae, Versailles.

Kupiec, J.J. (2014). Cell differentiation is a stochastic process subjected to natural selection. *In:* A. Minelli and T. Pradeu (eds), *Toward a theory of Development* (p. 155-173). Oxford University Press. Oxford. DOI:10.1093/acprof:oso/9780199671427.003.0010

Kupiec, J.J. and Sonigo, P. (2000). *Ni Dieu ni gène*. Seuil, Paris.

Laland, K.N. et al. (2015). The extended evolutionary synthesis: its structure, assumptions and predictions. *Proc. R. Soc. B,* 282: 20151019. http://dx.doi.org/10.1098/rspb.2015.1019

Lecointre, G. (1997). Dialogue Généticiens/Systématiciens : une culture du quantitatif contre celle du qualitatif ? *Biosystema* 15: 7-20.

Lecointre, G. (2015a). Descent (filiation). *In:* T. Heams, P. Huneman, G. Lecointre, and M. Silberstein (eds), *Handbook of Evolutionary Thinking in the Sciences* (p. 159-207). Springer, Dordrecht, Heidelberg, New-York, London.

Lecointre, G. (2015b). *Descendons-nous de Darwin ?* Le Pommier, Paris.

Lecointre, G. (2015c). La grande invention théorique de Darwin. *In :* G. Lecointre and P. Tort (eds), *Le Monde de Darwin* (p. 62-75). La Martinière, Paris.

Lecointre, G. (2017). Les contenants et les contenus. Que faire des invariants en Biologie ? *In:* T. Gaudin, D. Lacroix, M.C. Maurel, J.C. Pomerol (eds), *Sciences de la vie, sciences de l'information* (p. 129-143). ISTE Éditions, Paris.

Lecointre, G., Schnell, N.K. and Teletchea, F. (2020). Hierarchical analysis of ontogenetic time to describe heterochrony and taxonomy of developmental stages. *Sci Rep* 10, 19732. https://doi.org/10.1038/s41598-020-76270-4

Levin, M., et al. (2016). The mid-developmental transition and the evolution of animal body plans. *Nature* 531(7596), 637-641. doi: 10.1038/nature16994.

Lwoff, A. (1969). *L'ordre biologique*. Laffont, Paris.

Mayr, E. (1961). Cause and effect in biology. *Science* 134: 1501-1506. PMID 14471768.

Mayr, E. (1984). The triumph of the evolutionary synthesis. *Times Lit.* 2, 1261–1262.

Minelli, A. (2014). Developmental disparity. *In:* A. Minelli and T. Pradeu (eds), *Toward a theory of Development* (p. 227-245). Oxford University Press. Oxford. DOI:10.1093/acprof:oso/9780199671427.003.0015.

Moczek, A.P., (2012). The nature of nurture and the future of evodevo: toward a comprehensive theory of developmental evolution. *Integrative Comp. Biol.* 52, 108–119. doi:10.1093/icb/ics048.

Moczek, A.P. (2014). Towards a theory of development through a theory of developmental evolution. *In:* A. Minelli and T. Pradeu (eds), *Toward a theory of Development* (p. 218-226). Oxford University Press. Oxford. https://doi.org/10.1093/acprof:oso/9780199671427.003.0014.

Moreau de Maupertuis, P.L. (1751). Essai de Cosmologie. Leyde. https://gallica.bnf.fr/ark:/12148/bpt6k15100153?rk=21459;2

Müller, G. B. (2007). Evo- devo: Extending the evolutionary synthesis. *Nature Reviews Genetics*, 8, 943– 949.

Nelson, P. and Masel, J. (2017). Intercellular competition and the inevitability of multicellular aging. *Proc. Natl. Acad. Sci. USA* 114, 12982–12987. https ://doi.org/10.1073/pnas.16188 54114.

Nijhout, H.F. (1990). Metaphors and the role of genes in development. *Bioessays* 12(9), 441-6. doi: 10.1002/bies.950120908. PMID: 1979486.

Nowell, P.C. (1976). The clonal evolution of tumor cell populations. *Science* 194, 23–28.

Peluffo, A.E. (2015). The « Genetic Program »: Behind the Genesis of an Influential Metaphor. *Genetics* 200, 685-696.

Philippidou, P. and Dasen, J.S. (2013). Hox genes: choreographers in neural development, architects of circuit organization. *Neuron* 80(1), 12-34. doi: 10.1016/j.neuron.2013.09.020.

Raj, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135(2): 216–226.

Roux, W. (1881). *Der Kampf der Teile im Organismus*. Wilhelm Engelmann.

Roux, W. (2013). *La lutte des parties dans l'organisme* (T. Heams, ed., L. Cohort, S. Danizet-Bechet, A.L. Pasco-Saligny and C. Thébault, trad.). Éditions Matériologiques, Paris.

Schmid-Siegert, E. et al. (2017). Low number of fixed somatic mutations in a long-lived oak tree. *Nat. Plants* 3(12), 926-929. https://doi.org/10.1038/s41477-017-0066-9 (2017).

Schrödinger, E. (1944). *Qu'est-ce que la vie ?* Éditions du Seuil, coll. « Points Sciences », Paris, (1993).

Sonnenschein, C, and Soto, A.M. (1999). *The society of cells: cancer and control of cell proliferation.* Springer Verlag, New York.

Soto, A.M., Sonnenschein, C. and Miquel, P.A. (2008). On physicalism and downward causation in developmental and cancer biology. *Acta Biotheor* 56, 257-74.

Soto, A.M. and Sonnenschein, C. (2005). Emergentism as a default: cancer as a problem of tissue organization. *J Biosci* 30, 103-18.

Thomas, F. (2019). *L'abominable secret du cancer*. Humensciences, Paris.

Turajlic S. and Swanton, C. (2016). Metastasis as an evolutionary process. *Science* 352 (6282), 169-175. https ://doi.org/10.1126/science.aaf2784.

Vignaux, G. (1977). Les arguments pour une nouvelle « logique du vivant ». *In:* P. Achard, A. Chauvenet, E. Lage, F. Lentin, P. Nève, G. Vignaux (eds), *Discours biologique et ordre social* (p. 14-94). Coll. Science ouverte, Seuil, Paris.

Wagner, D. E. et al. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* 360, 981–987.

Watson, J. D. et al. (1987). *Molecular Biology of the Gene* (3rd edition). Cummings, Menlo Park, California.

Wilkins, A. S. (2013). The "genetic tool-kit": the life-history of an important metaphor. *In:* J. Streelman (ed.), *Advances in Evolutionary Developmental Biology* (p. 1–12). Wiley, Hoboken.

Willyard, C. (2016). Cancer: an evolving threat. *Nature* 532, 166-168. https ://doi.org/10.1038/53216 6a.

Woltering, J.M. et al. (2018). The skeletal ontogeny of *Astatotilapia burtoni* - a direct-developing model system for the evolution and development of the teleost body plan. *BMC Dev Biol.* 18(1):8. doi: 10.1186/s12861-018-0166-4. PMID: 29614958; PMCID: PMC5883283.

Zhao, Z.M. et al. (2016). Early and multiple origins of metastasic lineages within primary tumors. *Proc. Natl. Acad. Sci. USA* 113 (8), 2140–2145. https ://doi.org/10.1073/pnas.15256 77113 (2016).

Chapter 20

The Generalized Selective Environment

Hugh Desmond[1]

Abstract

In discussions about how Darwinism should be generalized, the concept of the environment has received relatively little attention. In this chapter I propose that the main explanatory work done by the "generalized selective environment" lies in introducing a contrastive explanation with agential explanations. This contrast comes in degrees. Keeping this contrast in mind helps disentangle some sources of confusion when the theory of natural selection is applied to humanities and social sciences, since these domains deal in entities that might otherwise be explained as products of human agency (e.g., human behavior, moral norms, scientific theories, social institutions).

Keywords: Natural selection – Human Agency – Cultural Evolution – Memetics – Evolutionary Psychology

---

[1] [1] Institut d'Histoire et de Philosophie des Sciences et des Techniques (IHPST), University of Paris I Panthéon-Sorbonne, Paris; Department of Philosophy, University of Antwerp.

20.1. Introduction

In biological context, nobody contests speaking of “the selective environment” of a population for explanatory reasons. One could disagree about how *real* the environment is, and to what extent all there *is,* is a profusion of ontologically disunified processes impacting individual organisms in different ways, ranging from the physiological to the ecological. However, endorsing an selectionist explanation of some trait – and despite all the debates about the relative importance of natural selection, there is no serious debate that natural selection occurs at least sometimes – implies some concept of “the selective environment” to which the trait in question has adapted.

By contrast, even this minimal endorsement of the environment concept is not obvious – like other fundamental concepts such as fitness – once the principle of natural selection is applied outside of biology. Scientific ideas, moral norms, or corporations are sometimes said to be “adaptive” or “selected for”. *To what* “environment” do they adapt, and what is the environment that defines their fitness?

The fundamental reason why this question is much more difficult in context of generalized evolution, is that many of the most active evolutionary subfields (especially cultural evolutionary theory: see introduction of this volume) target phenomena that are *also* products of human cognition and behavior, and which *also* interact in complex ways with social trends and social norms at the scale of communities. To explain why cetaceans evolved in the direction of limbless streamlined bodies, it seems very plausible to refer to a mind-independent reality, namely the aqueous environment. Products of human cognition are not like that, since often they cannot be entirely explained through a process of adaptation to the physical environment, and they are both shaped and shape human minds and human communities.

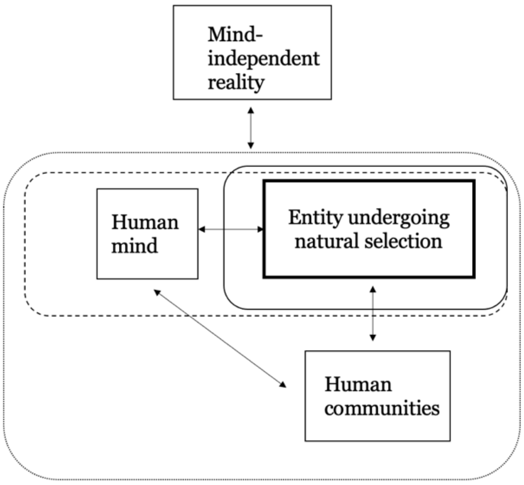


Figure 2: Where should the boundary between the selective environment and the (extended) evolving entity be drawn?

As a concrete illustration, consider one of the first evolutionary approaches to science, that of David Hull (Hull [1988] 2010). Hull proposed a memetic analysis, where scientific ideas (hypotheses, methodologies, epistemic values, etc.) played the role of replicators and books, journals, and scientists that of interactors.[2] Ideas shape the brains of scientists; the locutions of scientists in turn determined how successfully the idea could spread to other scientists. What, in this vision of the evolution of science, is the "environment" in which selection takes place?

One could think that it is the *minds* of individual scientists, as the "space" that ideas can "colonize". After all, memes were originally presented as "parasitizing" human minds (see Dawkins [1976] 2006). However, Dawkins seems to have primarily fashions, hypes, or ideologies in mind (see Dawkins [1976] 2006, 192 ff.). It is implausible to think the success of,

---

[2] In his words: "elements of the substantive content of science-beliefs about the goals of science, proper ways to go about realizing these goals, problems and their possible solutions, modes of representation, accumulated data, and so on." (Hull [1988] 2010, 434). Dawkins himself suggests that the true replicator is a "unit of information residing in the brain" (Dawkins 1982, 109), with words, music, images etc. the interactor.

for instance, Newtonian dynamics is entirely similar to the changing fashions regarding, for instance, beard length. While science is not immune to such dynamics -- the psychological appeal of a scientific theory (e.g., simplicity, elegance) plays its part in theory selection -- an idea must also generate empirical success in order to replicate successfully in a scientific community. In other words, mind-independent reality also determines the (long-term) fitness of scientific ideas, not just whatever biases or predispositions scientists may have.

Alternatively, one could broaden the ontological circle and think of the "selective environment" as consisting of features of mind-independent reality. In this line of thought, the scientist would assume the role of the extended phenotype, through which scientific ideas would interact with each other and with the environment. However, not only is this ontologically bizarre – dams are constructed by beavers, but neither scientists nor their behaviors are constructed by scientific ideas – but it implies a naïve falsificationist view of science. Scientists can keep on supporting ideas even if they seem "maladaptive to" (i.e., falsified by) the empirical environment (the Duhem-Quine thesis). In some cases this illustrates how scientists are agents with their own preferences and a capacity to intervene on the course of science. In other cases, in can illustrate how scientists can be irrational creatures with minds that can be "infected" by maladaptive scientific ideas.

Is the human mind part of the selective environment, or should the selective environment be placed outside of the human mind? When phrased at this higher level of abstraction, it becomes clear that we run into similar problems if we would enquire about the selective environment of corporations and social entities, moral norms, technological know-how, or languages. Thinking of the human mind as a selective environment is explanatory towards some types of phenomena; thinking of empirical reality as the selective environment is explanatory for other phenomena. It does not seem fruitful to argue about what is "the" selective environment, even for a very limited type of generalization of natural selection (i.e., memetic theory).

The natural temptation here is to adopt a skeptical stance, and see these conceptual problems as a confirmation of too much enthusiasm and too little rigor in generalizing Darwinism. Undoubtedly one should be careful what precisely are the explanatory ambitions of generalized Darwinism (see part III). However, insofar as a domain such as cultural evolutionary theory is explanatorily fruitful (following Boyd and Richerson 1985), one cannot judge that the "generalized selective environment" is simply a misleading metaphor. If there is such a thing as cultural selection, there *must* be a concept of a "cultural selective environment". This chapter aims at sketching the contours of the "generalized selective environment", and most attention will be given to the cultural selective environment as this is the most active area where generalized versions of natural selection are being applied.

My proposal consists of a negative and a constructive part. The negative part is partial endorsement of the skeptical stance towards generalized Darwinism: there is no fact of the matter what "the" selective environment is. It is not fruitful to look the causal or physical features of the world that actually determine selection pressures. There are no *truthmakers* of the selective environment in generalized settings. I survey two class of reason to support this skepticism, both of which are generalizable. First, environments are heterogeneous, and selective environments may sometimes be objectively vague. Second, individual and environment engage in reciprocal interactions, and thus the individual can act as the "environment of the environment", namely, as a cause of changes in selection pressures.

However, my endorsement of skepticism towards the generalized selective environment is only partial. If one shifts the question from trying to pin down truthmakers towards identifying *explanatory roles,* then a more general account becomes possible. The question to ask is, is there some generalized explanatory function that the selective environment plays across evolutionary contexts, whether biological, cultural, economic, scientific? If one surveys the different explanatory roles that the selective environment plays across contexts, are there any common features?

The account that this chapter outlines is that the generalized explanatory function is defined in contrast to *agency.* In particular, the generalized explanatory function of the selective environment is to *screen off* the explanatory role of agency in explanations of adaptation. This function can be realized in different ways, depending on what agency refers to. Agency can refer to non-human organismic agency, human agency, and divine agency. The focus in this paper will be mainly on how natural selection screens off human agency. This can be done in many ways and to varying degrees, and disambiguating between these ways and degrees helps clarify the distinctions between types of generalization of natural selection.

The paper is structured as follows. In the next section I will review the reasons why it is not fruitful to think of the selective environment as part of the furniture of the world, but that we should think of it as an epistemic or explanatory category instead. I then (section 3) will focus on four types of explanation of changes of honor norms – evolutionary psychology, memetics, dual inheritance theory, and cultural niche construction – and will identify the selective environment concept present in each, and contrast it with agential explanations (section 4). Based on this analysis of the selective environment of a moral norm, I will argue how it can be extrapolated and will present an account of the generalized selective environment (section 5).

20.2. Ontic and Explanatory Approaches to the Environment

The concept of the "environment" of course plays a large role in evolutionary theory, and an essential role in the theory of natural selection in particular. It is also traditional to acknowledge that the environment has rarely been explicitly problematized (From Brandon 1990 and Abrams 2009; to Huneman 2021 and Walsh 2021). Efforts to pin down the meaning of fitness, individuality, or heritability have all spawned large literatures, but the environment concept has not invited systematic reflection to the same extent.

The concept of environment has creeped into evolutionary thinking almost unnoticeably. Yet naturalists, including Darwin in *On the Origin of Species,* originally only spoke of a concrete plurality of "circumstances" impinging on the organism. It was Spencer who primarily popularized the notion in the English-speaking scientific community through his *Principles of Psychology* (Spencer 1855) – and the term either traces back to the translation of Comte's term of *le milieu* or Goethe's term of *Umgebung* (see Pearce 2010, 248). Darwin only started using the term from 1875 on, without any explicit reasoning being given for the adoption (Darwin 1875; Pearce 2010, 249). It seems like, at some moment, the explanatory need was felt to refer to "the environment" as abstract shorthand for the concrete plurality of circumstances.

If one attempts to reconstruct how precisely this abstraction of the circumstances occurs, one tends to find ontological disunity. After all, the environment concept is parasitical on the individuality concept: the environment contains everything that is "not" the individual. So if one defines individuals by their outer physical membranes or surfaces, the "environment" then contains all processes or features that are spatially outside these surfaces. If one defines individuals by what their immune systems accepts (Pradeu 2012), then the "environment" contains all processes or features that rejected by the immune system, including features that may be physically inside the individual (e.g. invading pathogens). If one defines individuals as Gibsonian agents, then the environment becomes a landscape of affordances that is unique to each individual (Walsh 2015). And crucially for our purposes, if one defines individuals as units of selection, the associated environment concept is the "selective environment" (Brandon 1990).

What is the selective environment? In textbook representations of natural selection, this is often glossed over. There the selective environment is typically represented by a single variable that can be toggled between two values, such as the "presence" or "absence" of sooted trees (Ridley 2004; Futuyma and Kirkpatrick 2017). In reality, say for a field biologist trying to determine the selective environment of a population, it may be genuinely unclear what processes or features contribute to selection, due to various forms of environmental heterogeneity. What could the metaphysical philosopher of biology say to the field biologist? What are the types of process that should be included in their determination of the selective environment? What are the truthmakers of the selective environment?

One problem here is that environmental heterogeneity, despite being ignored by typical textbook representations of natural selection, is the default state of natural environments. If an organism is mobile, their immediate circumstances change as they move about. Perhaps this variability can be discounted when the spatial range of organism – and by extension, of the population – is more or less well-defined. The biologist may then assign a "habitat" to the organism or population, where the habitat consists of with a number of "expected" processes or features that can go into defining the selective environment. However, in general, environments are continually changing due to exogeneous factors, including geological processes or climatological processes. They may change as individuals radiate into new habitats, for instance due to intragroup competition. They may change as rival species evolve, with new patterns of predation or symbiosis occurring.

In the abstract, it seems plausible to simply conclude that there is no "universal rule" for determining what processes or features constitute the selective environment of an organism (or population). A well-defined pattern of variability is necessary to be able to define the selective environment, and not all variation allows for a well-defined pattern to be inferred (this is argued in more detail in Desmond 2021, pp. 5-8). There is always a fallible epistemic leap from observations of recurrent phenomena to the probability with which they occur[3].

This does not imply some type of subjectivism, where the judgments of field biologists would merely reflect their arbitrary explanatory interests. Field biologists can still determine *in particular instances* what the selective environment is, but need specific information about life-history, salient ecological features, patterns of migration, and so on. For instance, stable patterns, or approaches to equilibrium, could be taken as evidence of a well-defined selective environment (for more detail on this, see Desmond 2018). The absence of a universal formula for infallibly judging whether a process is part of the selective environment does not imply that determinations of the selective environment are *merely* dependent on arbitrary explanatory interests.

A second conceptual roadblock in the way of attempts to pin down the selective environment, is that boundaries between individual and environment can become objectively blurred once reciprocal interactions take place. An organism may behave – through niche construction – in ways that alter the selection pressures on itself and its kin: the organism is the "environment of the environment". In such cases, to fully explain how organisms evolve through natural selection, one cannot simply refer to processes in the selective environment as if these are wholly external to the organism. One needs to refer also to how the behavior, development, or even metabolism of organisms impact their environment. The selective

---

[3] On this point, I am in agreement with (Abrams 2009; Bourrat 2020). A once-off occurrence (lightning strike) may significantly impact the expected number of offspring, but if the occurrence is truly once-off, there will be no way of verifying what exactly the probabilities are underlying the expectation number. This is why, when a freak environmental occurrence significantly alters reproductive numbers in an otherwise relatively stable environment, we tend to speak of "drift" instead of "selection".

environment thus becomes an intermediary in the causal link between organismic behavior and evolution.[4]

At what point does the environment-organism distinction break down? The concept of extended phenotype shows how it may break down with regard to certain causal features in the environment: the dam that the beaver builds can be seen as the beaver modifying its environment, but also as the genotype of the beaver expresses itself through a particular phenotype (Dawkins 1982). It may be difficult to choose whether dam should be viewed as an extension of the organism, or as part of the environment that affects the beaver. it is both. However, once the first generation has constructed the dam, it becomes an extrinsic variable for subsequent generations. The dam may continue to exist regardless of the behavior of subsequent generations of beaver. Hence for subsequent generations of beaver, for whom the dam predates them, the parts of the dam that perdure can more appropriately judged as being part of the selective environment rather than as extensions of the organism.

Reciprocal interactions such as these can knaw away at the fruitfulness of attempts to pin down the selective environment. As environment and individual become entangled – and the phenomenon of niche construction illustrates how this happens – it becomes increasingly hopeless to try to pinpoint what precisely belongs to the unit of selection and what belongs to the selective environment. This entanglement, however, is not very extreme in the case of the beaver and the dam, because the dam's continued existence does not depend on the beaver's continued existence. That is different with scientific theories, moral norms, or corporations. Their continued existence *does* depend on human's continued existence. That is why the entanglement between such cultural units of selection and the cultural selective environment is much more extreme. The next section will illustrate how this entanglement seems to preclude attempts to pin down the truthmakers of the cultural selective environment.

---

[4] This is the underlying rationale for viewing niche construction as an *agential process*, since not any alteration of the selective environment is niche construction, but only alterations that are a goal of the organism itself. This view is developed to some extent in (Aaby and Desmond 2021).

20.3. The Environment in Evolution of Honor Cultures

As a case study, I will focus on the cultural selection of honor norms, building on the work of Stefan Linquist (Linquist 2016) who teases apart the diverging predictive implications of different models of cultural evolution. Identifying how these models generate divergent predictions forces clarity on two core questions: what precisely are the units of selection, and what types of ontological factors determine the selective environment.

What are honor cultures? They describe communities characterized by norms that endorse violence as a morally justified response to insults, sleights, or other reputation-lowering actions. They appear to be more common among pastoral communities than among horticultural communities, and hence the question arises whether honor cultures are somehow adaptive to the types of circumstance faced by pastoral community.

The general adaptationist rationale for honor cultures, is that in some “socioecological environments” (Linquist 2016, 215) it will be adaptive to have a low threshold for violent retaliation and escalation against reputation-lowering behavior. Reputation-lowering action can snowball into theft, conflict, and exclusion of the insulted person. The reactive sensitivity to reputational threats (e.g., the disposition to react violently to insults) as a signal to others that entering into conflict will be costly. Honor norms postulate such sensitivity as the norm. Individuals can furthermore adopt certain behaviors (e.g., threats of retaliation) that signal to others that they adhere to such honor norms.

In general, such socioecological environments in which honor norms are adaptive are characterized by two variables (Linquist 2016, 216): whether or not individuals depend resources that can be easily stolen, and whether individuals can rely on a legal or social corrective mechanism to address perceived injustices. Both conditions tend to be met in pastoral communities. The main resource is livestock, which is more easily stolen than, for instance, housing or agricultural land. Moreover, these communities tend to be spread out geographically (to allow for grazing), which means most will live too far from any type of communal norm enforcement to be able to rely on the latter.

20.3.1. Four types of selective cultural environment

This is the *general* adaptationist rationale, but as such is underdescribed. What are the units of selection and the selective environment in this story? Linquist introduces the notion of the "socioecological environment" and distinguishes between two types of *evolutionary* unit: psychological phenotypes (patterns of behavior that characterize an individual) and cultural phenotypes (the norms of a community). Which unit is also the unit of selection depends on the model of cultural evolution employed. The following starts from his analysis of four models of cultural evolution – evolutionary psychology, memetics, dual inheritance theory, and cultural niche construction – and zooms in on the question how precisely each model conceptualizes the selective environment.

First, in the model of memetics, cultural variants are virus-like units hosted by human minds. The paradigmatic type of memetic evolution is the spread of a catchy tune, or the rise and fall of fashions. Here the unit of selection is the cultural phenotype, and the selective environment is the human mind. More sophisticated construals of memetics are possible, but they stretch the original conceptual fabric and are ultimately less satisfactory.[5] The units of selection spread by appealing to psychological dispositions of individual human beings.

For this model to successfully explain the evolution of honor norms, one pattern the data would need to show is that the spread of honor norms to be relatively independent of varying

---

[5] For instance, instead of interpreting the human mind as the environment for the meme, one could potentially view it as the extended phenotype of the meme. In this view, the selective environment would consist of the social and ecological conditions acting on individual humans. However, in this alternative construal, memetics no longer presents a distinctive model to e.g. dual inheritance theory, where human organisms are the main unit of selection and where cultural variants are understood as a fitness-enhancing trait. Second, this compromises memetics' explanatory simplicity for the core cases of fashion trends or catchy tunes – and hence brings memetics further away from why it got uptake in the first place. When Dawkins introduced the concept, he cites approvingly the gloss his colleague N. K. Humphrey gave of it: "When you plant a fertile meme in my mind you literally parasitize my brain, turning it into a vehicle for the meme's propagation in just the way that a virus may parasitize the genetic mechanism of a host cell." (Dawkins 1976, p. 192) It's also the construal of memetics that was later promoted enthusiastically (Blackmore 1999). So with these caveats in mind, it is fair to restrict memetics to the view where the human mind is exploited or harvested by memes in much the same way organisms may exploit the physical environment.

the fitness benefit or cost to individuals. As circumstances change, but the makeup of the human mind presumably does not[6], the memetic model predicts that honor norms would be able to continue to hijack human psychology with success. (This is not borne out by the data.)

The second model is that of evolutionary psychology, or more precisely, canonical evolutionary psychology (see Nettle, this volume). Here the human mind is hypothesized to have evolved through cognitive modules being selected for in the "environment of evolutionary adaptedness" (EEA), which is, loosely, the types of general conditions that are hypothesized to characterize ancestral hunter-gatherer societies. On this view, cognitive evolution tracks genetic evolution, and moreover, cognitive modules develop in a canalized fashion and thus are relatively insensitive to changing physical or social conditions. It does not imply that habits of behavior or of preference will develop uniformly among individuals: cognitive modules will express themselves different in different circumstances. However, the modules themselves are assumed to be rigid. The units of selection are genotypes, and the selective environment consists of the social and ecological factors that characterized the EEA (small communities; subsistence through hunting of game and the gathering of other forms of nutrition).

On this model of cultural evolution, one would expect to see honor norms as a quasi-universal cultural pattern: patterns of behavior that are robustly manifested despite cultural or ecological variation.[7] These are the predictions that are born out, with some plausibility (though not without controversy) with regards to mating preferences and strategies that differ (on a population level) between male and female genders and with regards to perception and reactions

---

[6] Again, one could attempt a more sophisticated interpretation of memetics here, where the makeup of the human mind can change in response to change in the environment. Human minds are, after all, highly plastic and can develop differently (in some respects) in different cultures. This interpretation of memetics would account for how some variants find more "fertile soil" in some cultures rather than others. However, this interpretation runs into the same problems as the interpretation above.

[7] Again, also evolutionary psychology could be interpreted in a more sophisticated fashion, where some plasticity would be allowed for. Thus the expression of cognitive modules might modulate in response to varying environmental input. However, the more plasticity is allowed, the weaker the distinctive explanatory power of evolutionary psychology is, and the more one needs to refer to concepts such as social and individual learning which feature more centrally in gene-culture inheritance theory and cultural niche construction theory.

to social status and reputation (Buss 2019). Hence one would expect similar patterns with regard to the distribution of honor norms and behaviors. (This is not borne out by the data.)

The third model is gene-culture dual inheritance theory (DIT). The paradigmatic example here is the spread lactose tolerance, which both involved the spread of know-how of animal husbandry as well as the spread of the alleles responsible for lifelong to digest lactose. There are two units of selection that propagate through two forms of inheritance: genotypes through genetic inheritance (i.e., biological reproduction) and cultural variants through cultural inheritance (i.e., social learning). In contrast to memetics, the cultural variant is not conceptualized as the unit of selection; rather, the variant is a property that affects the fitness of the *biological* unit of selection. Depending on one's view of natural selection this may be the individual or the genotype.

The fourth model is cultural niche construction theory (CNCT), where the idea is that cultural variants allow humans to create both social and ecological niches. A skill such as knowing how to construct artificial shelter allows for humans to carve out new ecological niche in colder climates. A skill such as shamanism allows individuals to carve out social niches in communities with division of labor. Like the previous model, the units of selection consist of both cultural variants and genotypes.

How do they differ? I propose (and this is where I depart more meaningfully from Linquist's presentation) that they differ in their conceptualization of the selective environment. In CNCT, the selective environment has an explicitly cultural character: behavior adapts to the norms in the social environment, and one norm can evolve to adapt to other, more established norms. By contrast, in DIT, the selective environment does not have this cultural component. The environment that determines the coevolution of genes for lactose tolerance and animal husbandry behaviors is an ecological environment: whether the environment can support cattle. In general, DIT is posited to explain how humans have been able to adapt to environments that change on short temporal and spatial scales (Boyd and Richerson 1985, chapter 4).

This account for the difference between CNCT and DIT, I believe, consistent with Linquists' and Sterleny's accounts, though perhaps more fundamental. On Linquist's interpretation, the key difference lies in the degree of functional integration of the cultural variants. In cultural niche construction, cultural variants form interlocking complexes or “traditions” to which individual human beings then adapt. Because humans adapt to the complex whole, a change in one of the elements of the tradition may not produce any material change in behavior. Honor cultures thus are characterized by reputation-maintenance norms (views on what types of response to behavior or threat may be considered “normal”), gender norms (views on what “normal” female and male behaviors are), and educational norms (what views can be passed on to the next generation). All of these tend to coincide, but on the cultural niche construction theory, a variation in, say, gender norms might not produce material change in behavior. By contrast, in dual inheritance theory, cultural variants evolve as autonomous units, and have a more one-to-one relation to behavioral patterns.

In my account here, functional integration is given a different interpretation: it is what ensures that cultural complexes become decoupled from individual behavior, in such a way that cultural complexes become part of the selective environment for individual humans. In NCT, there is something as the “cultural selective environment”, which is precisely made possible by the fact that cultural variants do not evolve as autonomous units in response to ecological change, but instead interlock and decouple both from individual behavior and ecological change.

Sterelny accounts for the difference in terms of “cultural scaffolding”, where the transmission of some cultural variants cannot be done simply through imitation of the whole, but require guided learning where the cultural variant is broken down into components and which then are imitated. For instance, manufacturing stone hand axes is a process composed of multiple steps, each of which are taught separately. Thus, for the cultural variant of stone hand axes to be transmitted, and the associated niche it allows for, there must be a teacher that acts as a scaffolding for this transmission process. (Once the transmission is complete, the scaffolding can be removed.) However, also the process of stone hand axe construction can be viewed as depending on a complex whole consisting of component skills, where the whole has functional value but only the component skills are transmissible through imitation. The learning process set up by the teacher (i.e., the scaffolding) forms a “cultural selective environment”, to which the learner must adapt (and which the learner does not construct, or affect through their own activities) and which follows a rate of change that is decoupled from change in the ecological environment.

Table 1 summarizes this discussion of the four models of cultural selection. The explanandum target by all four models is the observed pattern of human behavior across cultures, but different explanantia are posited. In particular, different types of environment are posited to play the role of “selective environment”.

Model	**Explanans**				**Reciprocal interactions**
	Unit of selection	**Mode of Inheritance**	**Selective environment**		
Memetic	Cultural variants	Social learning	Human minds (preferences of various kinds: biased cognitive mechanisms, prejudices, needs, desires)		No
Evolutionary psychology	Genotypes (esp. those predisposing the canalized development of certain cognitive programs)	Genetic	Ancestral socio-ecological environment (esp. that of hunter-gatherer societies: the EEA)		No
Dual inheritance theory	Cultural variants and Genotypes/ Individuals*	Social learning and genetic	Ecological environment		No
Cultural niche construction	Cultural variants and Genotypes/ Individuals*	Social learning and genetic	Cultural environment (proximal)	Ecological environment (distal)	Yes

***Table 1:* Explanans categories across cultural evolution models.** The explanandum for all four models is the same: patterns of human behavior, both across as well as within cultures.

* The question whether genes or organisms are the unit of selection is not relevant here.

20.3.2. Ontological disunity

What lesson does this case study hold for how the selective environment should be generalized? It is that, in generalized settings, it is not only difficult to pin down the components of the selective environment with great exactness, but it does not even seem possible to pin down the *kind of thing* a selective environment is.

In biological settings, it can be difficult to pin down what processes or features of the environment can be considered as part of the selective environment when there is a lot environmental heterogeneity (Desmond 2021a); however, it is not in doubt that the selective environment possesses a spatial character. It refers to processes "outside of" different organisms that impact fitness differentials. However, in the context of cultural evolution, it seems undecided whether the selective environment has a spatial nature. In some models of cultural evolution, such as dual inheritance theory (or at least, according to the construal above), the selective environment has a relatively straightforward spatial character. By contrast, in memetics, the selective environment consists of an ensemble of cognitive mechanisms. The meme itself is not a tangible entity occupying a certain volume of space: Dawkins characterized it as a "unit of information residing in the brain" (Dawkins 1982, 109).

Whether a selective environment is made up of cognitive mechanisms, or of ecological processes, or of some ensemble of social norms, seems to mainly depend on explanatory purposes. In the specific case of evolution of honor norms, Linquist concludes that cultural niche construction theory best predicts the data given by the Human Relations Area File (Linquist 2016, 230–34). Evolutionary psychology is discounted by evidence that norms towards aggression do not always correlate with pastoral vs. horticulturalist conditions (Linquist 2016, 233–34). This suggests that there is no automatic cognitive module responsible for honor norms that is triggered by specific conditions. Dual inheritance theory, in turn, is discounted by the fact that specific norms towards aggression covaried with the behavior that parents expected or tolerated among children. This suggests that there is something as a "cultural environment", and not just distinct cultural variants, that is responsible for the spread and maintenance of honor norms.

For the explanation of honor norms, niche construction theory may offer the best explanation. However, for other explananda, niche construction theory may no longer be the best explanation. For instance, if one were to explain the dynamics of some fashion trend that seems to run counter to some aspect of normal biological functioning, such as the practise of foot binding, memetics might be a parsimonious explanatory model for that phenomenon. It would seem appropriate to generalize the principle of natural selection in such a way that the human mind appears as the selective environment, sifting through adaptive and less adaptive memes. Alternatively, if one were to explain why individuals attach great importance to social status – regardless of culture, gender, age, or personality – the model offered by (canonical) evolutionary psychology might offer a more appropriate explanation. It would seem appropriate to refer to the ancestral hunter-gatherer environment, in which cooperation was crucial for reproduction and survival, and where one needed to maintain esteem in the eyes of others in order to be part of collaborative networks.

Does the generalized selective environment consist of out of ecological conditions (just as the biological selective environment), or does it consist of cognitive modules, or of ensembles of social norms -- which are cognitive phenomena, but strangely non-reducible to the cognition of individuals? Asking what "thing" the generalized selective environment is, does not seem to lead enquiry down a fruitful path. If the objective-causal nature of natural selection and the selective environment can be doubted even in the context of biological evolution (Walsh 2000; Matthen and Ariew 2002), it at least seems *explanatory* to think of the selective environment as referring to the extra-organismic space consisting of the causal processes affecting natural selection. However, even this tenuous connection between ontology and explanatory value seems cut when it comes to the selective environment in generalized setting. Sometimes it may be useful to think of the human being as the unit of selection (evolutionary psychology or DIT); sometimes it may be useful to think of the human being as the environment (memetics). Sometimes it may be useful to think of cultural norms as the unit of selection (DIT); sometimes it may be useful to think of them as constituting the selective environment (NCT). The distinction between unit of selection and selective environment no longer seems to be “real” in any robust sense, in the way that the temperature, moisture, acidity in the garden can relatively unproblematically be said to constitute the “environment” of a plant.

20.4. The Contrast Class: Agential Explanations of Honor Culture

The preceding discussion could give credence to an instrumentalist reading of the generalized selective environment: one invokes a different environment concept according to the type of prediction one wishes to generate. The selective environment is a modelling tool, but not much more can be said of it. However, this type of approach opens up a more fruitful line of enquiry. Instead of asking what ontological features unify different conceptions of the selective environment, one can ask what *epistemic* or *explanatory* features are held in common. Why is the selective environment invoked at all? This is a question that does not come into focus if one is merely comparing different evolutionary models, with different associated understandings of the selective environment. The question to ask is, could the target explanandum not be explained *without* referring to units of selection and the selective environment?

My proposal is that the true contrast class of selectionist explanations are agential explanations. The true contrast class for explaining the evolution of honor norms through natural selection is explaining the same norms as the product of a reasonable deliberation about moral norms, or a lack thereof because of prejudice and bias.

This contrast class is often forgotten because of the intense competition between various models of cultural evolution. One could speculate that this competition is somehow maintained by disciplinary boundaries (cognitive niches?) in academia: cultural evolutionists discuss and debate other cultural evolutionists, whereas historians or cultural anthropologists discuss and debate other historians or cultural anthropologists. The contrast class may even be deliberately ignored, because, in contrasting selectionist explanations with rational explanations, it spans the is-ought divide. Nonetheless, in ignoring this contrast class, we lose sight of the real explanatory role played by the (generalized) selective environment.

What would an agential explanation of the evolution of honor look like? Such an explanation would appeal to the *intentions* and/or *reasons* humans have for adhering to or rejecting honor norms. Agential explanations may assume different forms, and I will discuss two: pragmatic and moral explanaitons.

In *pragmatic explanations,* the behavior is explained by (1) an agent acting to maximize benefit (however defined), and (2) the belief of the agent that the explanandum behavior maximizes benefit. In order to apply this to the case of honor norms, we can introduce some more detail from Nisbett and Cohen's classic account of the honor culture in the Southern United States. According to their account, violent responses (or threats thereof) to reputational threats can be understood as a (rational) strategy of deterrence of future threats. Thus, a pragmatic explanation refers to a process of adaptation to the social environment, but the process of adaptation is not explainable by natural selection alone. Rather, there are agents that perceive and understand certain challenges in their social environment, and choose their response accordingly. The agent may make mistakes, and wrongly calculate which strategy is the benefit-maximizing one. However, also here the core explanans is the agential reasoning process. No reference needs to be made to a selective environment.

Such pragmatic explanations are most familiar to evolutionary thinkers, as they are interpreted as the "maximizing agent analogy", a shorthand for a selectionist explanation (Sober 1998; Martens 2016; Okasha 2018). It remains a point of contention to what extent this analogy can be used for explanations of organismic behaviors (especially altruistic behaviors). It is even more doubtful whether human behavior can be explained by means of natural selection.

For instance, an evolutionary model such as that present in evolutionary psychology explains human behavior by referring to the triggering of automatic modules shaped in the ancestral environment. In this view, the reasoning process of the agent is a secondary phenomenon. This does not mean that the capacity for reason cannot be explained through a process of selection. For instance, it has been suggested that the capacity for reason is adaptive by (1) enabling coordination between agents in cooperative activities, (2) communicating beliefs even if the speaker lacks trustworthiness (Mercier and Sperber 2017). Reason and even agency itself could be the target of selectionist explanation. The point is that, giving a selectionist explanation of a phenomenon entails that an agential explanation is *not* given. One refers to the action of the selective environment, favoring certain modes of cognition or behavioral patterns over others. It is the environment that is doing the "choosing" and "selecting" rather than agent itself.

The contrast between selectionist and agential explanations is even clearer when it comes to moral explanations of behavior. *Moral explanations* invokes a different type of reasoning principle to pragmatic explanations: the agent acts not in order to maximize benefit, but in order to maximize some value that can be deemed intrinsically good (i.e., praiseworthy in themselves).[8] What would a moral explanation of honor culture look like? It would refer to honor as intrinsic good that can and should be pursued for its own sake, independently of wealth, suffering, or even happiness. As Dan Demetriou documents, this is exactly how individuals inhabiting honor cultures experience honor norms (Demetriou 2014). A whole honor morality could be construed, consisting of various norms, including the norm of fair play in honor competitions, or the norm that one should respond to challenges to one's honor (Demetriou 2014, 902). From the perspective of honor morality, the presence of honor cultures is explained as *morally right* – and thus hardly in need of an explanation.

Moral explanations can also explain behavior as an *absence* of moral reasons. Thus, if one adopts liberalism as the background moral framework, the value of human autonomy is prioritized over honor. The core norms in a liberal society are norms such as: the right to property or the right to pursue one's life as one sees fit (without harming others). From the perspective of a liberalism, honor cultures would simply appear to be *morally wrong*. This is also a type of moral explanation of honor cultures: honor cultures are explained as due to a lack of moral agency. The individuals in honor cultures are biased and prejudiced: this may be due to a lack of education, or due to material deprivation. Whatever the reason, the main explanans

---

[8] Given this way of putting the distinction between pragmatic and moral explanation, utilitarianism could be seen as an intersection between pragmatic and moral explanations ("benefit" is narrowly defined as pleasure and pain, which is an intrinsic good on hedonistic utilitarianism).

in such a moral explanation is the widespread presence of a curtailed or truncated agency, such that individuals are incapable of (morally) reasoning correctly.

Although it will not be further pursued here, this line of argumentation suggests where the tension between evolutionary (selectionist) and traditional rationalist perspectives on morality should be sought: not in issues about the existence or nonexistence of objective moral truth (Street 2006; Sterelny and Fraser 2017), but in the competition between two types of explanation, selectionist or agential, for status of the best explanation of behaviors which appear to driven by moral reasoning. After all, an agent may *believe* that their behavior is driven by moral reasoning, and still *not in fact* driven by moral reasoning. Their behavior could perhaps be better explained by cultures and norms that evolved to adapt to certain ecological conditions. Taking such a selectionist stance on the evolution of honor cultures seems to be relatively uncontroversial; where the question becomes much more controversial is when it concerns aspects of contemporary morality, and whether these may also be explained through recourse to natural selection.

In sum, in agential explanations, the main explanans is the agent and its process of reasoning or deliberation. This suggests a more general, yet more accurate way of characterizing the generalized selective environment, namely as an explanatory principle that screens off any need to appeal to agency.

20.5. Generalized Selective Environment

Based on these considerations, we can now propose an account of the generalized selective environment:

> The **generalized explanatory function** of the selective environment is to posit an explanans of adaptive evolution in such a way that agency is screened off.

This explanatory function can be realized in different ways: some selective environment concepts screen off the role of agency *entirely,* whereas others only do so partially. Let us first apply this account to the models of cultural evolution previously discussed. Then, I will suggest how the generalized selective environment can be understood even more generally, in contrast to three forms of agency: divine agency, organismic agency, and human agency.

20.5.1. The selective environment in various models of cultural evolution

What does the selective environment "do" in memetics? It serves to represent the human mind as a resource to be harvested, exploited, parasitized, invaded. In other words, not only is there no role for human agency, but the human mind is represented as an entirely passive quantity, like a natural resource to be mined. Any explanatory role that some reasoning process *could* play is screened off by the exploitable cognitive modules involved in that reasoning process. The *gestalt*-switch involved here is that an active reasoning process is presented as a process that is controlled by memes. This control by memes becomes the explanans that screens off the explanans of human agency. (The only problem is that this mode of explanation has only limited success, and in affirming the reality of moral agency, Dawkins himself seemed to have implicitly endorsed at least one limitation to the applicability of memetics (Dawkins [1976] 2006, 2).)

Evolutionary psychology seems to allow for the existence of human agency – reason itself, after all, may be the outcome of a process of selection (Mercier and Sperber 2017) – but again agency does not play any crucial explanatory role. Evolutionary psychology need not be used to defend extreme claims, such as that all human behavior is determined by cognitive modules, but in the explananda it is interested in, it does propose an explanation that screens off the role of human agency. For instance, in being attracted to certain properties in potential mates, men and women may believe they *themselves* are the source of this desire or preference. However, explanations in evolutionary psychology, by referring to a selective environment, will seek to screen off such agential explanations, and instead refer to cognitive modules for mate preference that were shaped in the ancestral environment.

Dual inheritance theory and cognitive niche construction theory also screen off agency as explanans, but do so in a more subtle way. In positing the importance of social learning, the implication is that a lot of human behavior is influenced by external cultural norms. How social learning is understood determines the explanatory role of agency. For instance, in DIT, the tendency is to model the spread of cultural variants (such as animal husbandry know-how) as a diffusive process: individuals imitate successful cultural variants, but there is no need to invoke real intentionality or choice.[9]

By contrast, in cultural niche construction, there may be a limited role for agency. If an individual learns to construct stone hand axes in a certain way, then this is partially due to the existence of a selective cultural environment, to which individuals have no choice by to adapt. In this respect, cultural niche construction screens off the explanatory role of human agency. However, during the social learning process, the individual cannot necessarily be adequately represented as a passive receptacle for the knowledge imparted by the teacher. For instance, the skill of hand axe construction needs to be divided into component parts, and while each these component skills could perhaps be learned by a process of imitation, the learner is still responsible themselves for the integration of the component skills. Nonetheless, cultural evolutionary theory, by explaining the adaptiveness of human behavior through natural selection (and selection by the environment), serves to undermine the idea that humans are autonomous, self-determining agents (this relation between cultural evolutionary theory and autonomy is explored further in Desmond 2021b).

20.5.2. Screening off Divine Agency

Screening off divine agency was, historically, the most important feat of Darwin's theory of natural selection. Darwin's contemporaries (e.g Paley) were concerned whether adaptive complexity in nature required divine agency as an explanans. They were aware that human agency could lead to impressive evolution within a species (dog breeding through artificial

---

[9] Of course, this statement is conditional on the underlying cognitive neuroscience of imitation (see Frith et al. 2003). Can imitation be fully explained in a basic mechanistic way, such as in terms of firing up mirror neurons? Or does one need to also refer to the capacity to create mental representations of the skill in order to successfully imitate it? In the latter case, some agential capacity creeps in.

selection). However, it was unclear how the adaptive complexity found across species could be explained through anything other than divine agency.

By contrast, with the concept of natural selection, there was no need to invoke a Paleyan God who designed organisms like a master watchmaker might design a watch. Against this background, a lot of the initial explanatory force of *On the Origin of Species* can be understood as charting an explanatory course that makes divine agency superfluous. Explanations by natural selection did not invoke agency, but rather what some philosophers have called population-thinking (Ariew 2008): a sufficiently long succession of a large number of insensible variations, some with a slight edge over others, that over time leads to large changes.

20.5.3. Screening off Organismic Agency

Invoking the selective environment can also function as a way to screen off organismic agency. Historically, this explanatory role positioned natural selection as rival to Lamarckian adaptation (where actions of the organism would determine evolution). The screening off of organismic agency has never been as conclusive as that of divine agency. Darwin himself seemed to hesitate. While it is relatively clear that Darwin did not believe in the necessity to invoke divine agency as an explanans, he is much more ambivalent about organismic agency. For instance, there are suggestive passages in the *Descent of Man* where he ascribes an evolutionary role to aesthetic preferences independently on their impacts on fitness (roughly: peacock tails are explained because peahens find them beautiful, and not because peahens have been shaped by natural selection to select for costly signals of fitness).

Today, it remains unclear to what extent organismic agency can be screened off as explanans in evolutionary explanations. Many of the developments in the extended synthesis, ranging from niche construction (Laland, Matthews, and Feldman 2016) to plasticity-first evolution (Levis and Pfennig 2016; West-Eberhard 2003), suggest that organisms may have an "active" role to play in explanations of adaptive evolution, in such a way that their role cannot simply be reduced to processes of natural selection.

Some variation occurs when the selective environment is defined by environmental heterogeneity. Phenotypic plasticity is adaptive to heterogeneous environments: however, the selective pressure for plasticity is only well-defined when the environmental patterns of heterogeneity are themselves well-defined (per the sense in (Desmond 2021, drawing on Levins 1968). Is such plasticity truly "agential"? One could debate whether one needs the concept of agency to describe, for instance, the elongation of a plant stem, and whether a developmental mechanism shaped by natural selection may suffice. Where the selective environment can less clearly screen off agency is when organisms and populations radiate into novel environments: these may be spatially novel habitats, but may also be new sources of nutrition that populations were able to access and exploit.

Thus, one could further distinguish between two broad explanatory roles the selective environment can play: the selective environment as "pseudo-agent" determining the fitness values of the organism, and the selective environment as repository of opportunities or resources ("affordances") for the organism to exploit. In the latter conception, the agency of the organism is emphasized (this is view in (Walsh 2015)).

The term "natural selection" indicates that the selective environment was primarily intended to adopt the role of a pseudo-agent. Lexically, "selection" is agential language, in the sense that selection denotes some type of goal. Otherwise Darwin could have talked about of "natural sifting" or "natural sorting" which – although they too admit of agential interpretations – are more easily interpreted as non-agential than "selection". "Natural sifting" would have been more in keeping with the statistical nature of natural selection (Matthen and Ariew 2002). Nonetheless, Darwin was attached to the term natural selection, despite Wallace purportedly preferred Spencer's phrase "survival of the fittest" because it was less agential (Lennox 2019). Darwin's own description of natural selection, rife with agential language (emphases mine):

> natural selection is daily and hourly *scrutinising*, throughout the world, every variation, even the slightest; *rejecting* that which is bad, *preserving* and adding up all that is good; silently and insensibly *working*… (Darwin [1859] 2008, 66)

Why did Darwin stick to the term “natural selection”? In one of his letter’s, he refers to two reasons: (1) the continuity with agential language in physics (“attractive” or “repulsive” forces; “laws” that “rule” the universe) or chemistry (preferences of acids), (2) a handy abbreviation of a much more complicated causal process:

> In the literal sense of the word, no doubt, natural selection is a false term; but who ever objected to chemists speaking of the elective affinities of the various elements? — and yet an acid cannot strictly be said to elect the base with which it in preference combines. It has been said that I speak of natural selection as an active power or Deity; but who objects to an author speaking of the attraction of gravity as ruling the movements of the planets? Every one knows what is meant and is implied by such metaphorical expressions; and they are almost necessary for brevity. So again it is difficult to avoid personifying the word Nature…(Peckham’s Variorum text, 164)

Nonetheless, it is has remained a point of contention to what extent organismic agency can be screened off entirely, throughout ecology and evolution. When the selective environment is conceptualized as a repository of opportunities, there is still some sense of the Darwinian “struggle for existence” or Spencerian “survival of the fittest”, even though the victors in that struggle may be those best able to exploit novelty and thus avoid direct competition with others. There is clearly a larger explanatory role here for positing the organism as causal origin of fitness-altering behavior (instead of the fitness-altering behavior being explainable in terms of an etiology of selection in previous generations). For purposes here, one could judge such a concept of the selective environment to only partially screen off organismic agency.

20.5.4. Screening off Human Agency

As the selective environment has been generalized to new domains, the screening off of human agency has become the most important explanatory dimension of the selective environment.

In many of the domains associated with generalized Darwinism, human agency plays a core explanatory role. The introduction to this volume lists: economics, anthropology, sociology, psychology, epistemology, ethics, aesthetics, literary studies, archaeology, medicine, history, computation, electronics, history of science, history of technology.

Why has there been less invasion of physics of chemistry by Darwinian explanatory schemes? If one analyzes generalizations of Darwinism in terms of the abstract criteria of variation, differential reproduction, and inheritance (e.g. Hodgson and Knudsen 2006), it may not seem clear why such criteria are taken up enthusiastically in comparative psychology but less so in solid-state physics. It is not that they cannot be applied: Quantum Darwinism uses natural selection to offer a resolution, or at least a frame of analysis, of the measurement problem. Why has the application of Darwinian explanatory schemes to psychology spawned a whole subfield in the way that Quantum Determinism has not? While I do not attempt a full answer here, pointing to the explanatory function of screening off (human) agency does help account for this: agency, whether human or divine, has not been standardly invoked for explaining

As with the screening off of organismic agency in evolutionary biology, it is likely that the screening off of human agency is far from complete in these domains of generalized Darwinism. In fact, one could expect that the screening off of human agency is more likely to largely fail, and meet only patchy success. The explanatory scheme of evolutionary psychology can be used to explain with reasonable success why we have certain mate preferences (even though it may seem otherwise irrational that males and females have such divergent mate preferences), by reference to the socio-ecological environment typical of small-scale hunter-gatherer communities. However, it sometimes is also used to claim that humans are unlikely be adaptive to contemporary environments, because “There was, for example, no police in the primeval savannas, nor were there planes or hijackers…”. This type of screening off of human agency – i.e., claiming that humans cannot optimally deal with airplanes because there were no airplanes in the EEA – is of course not very plausibles.

If one surveys the domains to which Darwinian modes of explanation are now being applied – a long but non-exhaustive list is given in the introduction to this volume – then one general pattern that seems to emerge is that some of the domains where Darwinian explanations make most inroads are those where agential explanations held the status-quo: psychology or anthropology.

Domain	**Explanandum**	**Classic Explanans (agent)**	**Darwinian explanans (selective environment)**
Economics	Formation of corporations Division of labor	Utility-maximizing agents	Markets
History	Salient events in human past	Human agents	Ecological or socio-ecological environment
Sociology	Structure of institutions	Group agency	Socio-ecological environment
Literary studies	Structure of a text	The intentions of the author; the norms and values of the author's contemporaries	Socio-ecological environment
Ethics	Endorsements certain actions as "good".	Rational agent deliberating about standards of good	Socio-ecological environment where such endorsements give some advantage (biological fitness, social status, power)
Aesthetics	Endorsements of certain objects as "beautiful"?	Rational agent deliberating about standards of beauty	Socio-ecological environment that associates those objects with an advantage (biological fitness, social status, power)
Epistemology	Claims to "know" certain propositions	Rational agent deliberating about	Socio-ecological environment where such claims give some

		evidence and prior beliefs	advantage (biological fitness, social status, power)
History of science/ technology	Patterns of adoption and rejection of ideas and technologies	Rational agents deliberating about the merits and demerits of ideas and technologies	The ensemble of mechanisms, biases, preferences characterizing human minds *OR* The socio-ecological environment where ideas and technologies may give benefits

In a second category of domains, the "evolutionary" approach consists of replacing the human designer by a selective environment. These evolutionary approaches do not *entirely* banish human agency, as the human agent is still expected to engineer the selective environment.

Domain	**Goal**	**Classic approach (agent)**	**Darwinian approach (environment)**
Evolutionary medicine	Health-promoting interventions on the human body	Medical scientist who understands physiological mechanisms, and can manipulate outcomes accordingly.	Set of ecological factors that favor certain physiological mechanisms over others
Evolutionary computation	Efficient algorithms	Computer scientist who designs an algorithm that produces a solution.	Set of incentives (or values) that favor some possible solutions over others.

20.6. Conclusion

One of the roadblocks preventing a fruitful analysis of selective environment in a generalized setting is that it is an ontologically disunified concept even within the relatively narrow context of biological evolution. Once one generalizes, sometimes the selective environment seems to be identified with the minds that cultural variants infect (memetics), other times with the socioecological environment of hunter-gatherers (evolutionary psychology), and yet other times with a collection of the social norms and ecological features impacting the biological fitness of individual humans (dual inheritance theory). Is there anything general that could be said? The proposal was that the generalized selective environment should be defined in terms of the *explanatory function* it is to fulfill, and this function is essentially contrastive: the selective environment supplants agency as the prime explanans. This account of the generalized selective environment captures very well what Darwin originally intended with his concept of natural selection (supplanting divine agency), and captures how the concept of natural selection was further developed in the Modern Synthesis (with the effect of supplanting organismic agency). Today, it is especially as a rival to human agential explanations that the theory of natural selection is attracting most attention.

However, the screening-off of human agency does not happen uniformly, and can be done in different ways for different phenomena. These ways are distinguished by: (1) the degree of screening off (can one safely ignore agency, or does it have a partial explanatory role to play), (2) whether the agent is identified with the selective environment, or becomes the mere expression of the unit of selection.

Acknowledging these distinctions can help avoid confusion associated with attempts to find what characterizes "*the* Darwinian" approach. While one could arguably identify a "pure" Darwinian approach with screening off agency entirely (i.e., doing to human agency what Darwin did to divine agency), in practice, fruitful Darwinian approaches in domains such as psychology or anthropology will only screen off agency partially, and will sometimes model humans as resources to be exploited, and sometimes as expressions of underlying cultural variants.

References

Aaby, Bendik Hellem, and Hugh Desmond. 2021. "Niche Construction and Teleology: Organisms as Agents and Contributors in Ecology, Development, and Evolution." *Biology & Philosophy* 36 (5): 47. https://doi.org/10.1007/s10539-021-09821-2.

Abrams, Marshall. 2009. "What Determines Biological Fitness? The Problem of the Reference Environment." *Synthese* 166 (1): 21–40. https://doi.org/10.1007/s11229-007-9255-9.

Ariew, André. 2008. "Population Thinking." In *The Oxford Handbook of Philosophy of Biology*, edited by Michael Ruse. https://doi.org/10.1093/oxfordhb/9780195182057.003.0004.

Blackmore, Susan. 1999. *The Meme Machine*. Oxford: Oxford University Press.

Bourrat, Pierrick. 2020. "Natural Selection and the Reference Grain Problem." *Studies in History and Philosophy of Science Part A* 80 (April): 1–8. https://doi.org/10.1016/j.shpsa.2019.03.003.

Boyd, Robert, and Peter Richerson. 1985. *Culture and the Evolutionary Process*. Chicago: The University of Chicago Press.

Brandon, Robert N. 1990. *Adaptation and Environment*. Princeton University Press.

Buss, David M. 2019. *Evolutionary Psychology: The New Science of the Mind*. 6th Edition. New York: Routledge.

Darwin, Charles. 1875. *The Variation of Animals and Plants Under Domestication*. 2nd ed. London: John Murray.

———. (1859) 2008. *On the Origin of Species*. Oxford World's Classics. New York: Oxford University Press.

Dawkins, Richard. 1982. *The Extended Phenotype: The Long Reach of the Gene*. Oxford University Press.

———. (1976) 2006. *The Selfish Gene*. Oxford University Press.

Demetriou, Dan. 2014. "What Should Realists Say About Honor Cultures?" *Ethical Theory and Moral Practice* 17 (5): 893–911. https://doi.org/10.1007/s10677-014-9490-3.

Desmond, Hugh. 2018. "Selection in a Complex World: Deriving Causality from Stable Equilibrium." *Erkenntnis* 83 (2): 265–86. https://doi.org/10.1007/s10670-017-9889-z.

———. 2021a. "Adapting to Environmental Heterogeneity: Selection and Radiation." *Biological Theory*, March. https://doi.org/10.1007/s13752-021-00373-y.

———. 2021b. "In Service to Others: A New Evolutionary Perspective on Human Enhancement." *Hastings Center Report* 51 (6): 33–43. https://doi.org/10.1002/hast.1305.

Frith, C.d., D.m. Wolpert, Andrew N. Meltzoff, and Jean Decety. 2003. "What Imitation Tells Us about Social Cognition: A Rapprochement between Developmental Psychology and Cognitive Neuroscience." *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358 (1431): 491–500. https://doi.org/10.1098/rstb.2002.1261.

Futuyma, Douglas J., and Mark Kirkpatrick. 2017. *Evolution*. Sunderland: Sinauer.

Hodgson, Geoffrey M., and Thorbjørn Knudsen. 2006. "Why We Need a Generalized Darwinism, and Why Generalized Darwinism Is Not Enough." *Journal of Economic Behavior & Organization* 61 (1): 1–19. https://doi.org/10.1016/j.jebo.2005.01.004.

Hull, David L. (1988) 2010. *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. University of Chicago Press.

Huneman, Philippe. 2021. "What Is It Like To Be an Environment? A Semantic and Epistemological Inquiry." *Biological Theory*, November. https://doi.org/10.1007/s13752-021-00390-x.

Laland, Kevin, Blake Matthews, and Marcus W. Feldman. 2016. "An Introduction to Niche Construction Theory." *Evolutionary Ecology* 30: 191–202. https://doi.org/10.1007/s10682-016-9821-z.

Lennox, James. 2019. "Darwinism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2019. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2019/entries/darwinism/.

Levins, Richard. 1968. *Evolution in Changing Environments: Some Theoretical Explorations*. Princeton, NJ: Princeton University Press.

Levis, Nicholas A., and David W. Pfennig. 2016. "Evaluating 'Plasticity-First' Evolution in Nature: Key Criteria and Empirical Approaches." *Trends in Ecology & Evolution* 31 (7): 563–74. https://doi.org/10.1016/j.tree.2016.03.012.

Linquist, Stefan. 2016. "Which Evolutionary Model Best Explains the Culture of Honour?" *Biology & Philosophy* 31 (2): 213–35. https://doi.org/10.1007/s10539-015-9515-x.

Martens, Johannes. 2016. "Inclusive Fitness and the Maximizing-Agent Analogy." *The British Journal for the Philosophy of Science*, June, axw003. https://doi.org/10.1093/bjps/axw003.

Matthen, Mohan, and André Ariew. 2002. "Two Ways of Thinking About Fitness and Natural Selection:" *Journal of Philosophy* 99 (2): 55–83. https://doi.org/10.2307/3655552.

Mercier, Hugo, and Dan Sperber. 2017. *The Enigma of Reason*. Cambridge, Massachusetts: Harvard University Press.

Okasha, Samir. 2018. *Agents and Goals in Evolution*. Oxford: Oxford University Press.

Pearce, Trevor. 2010. "From 'circumstances' to 'Environment': Herbert Spencer and the Origins of the Idea of Organism-Environment Interaction." *Studies in History and Philosophy of Biological and Biomedical Sciences* 41 (3): 241–52. https://doi.org/10.1016/j.shpsc.2010.07.003.

Pradeu, Thomas. 2012. *The Limits of the Self: Immunology and Biological Identity*. Oxford ; New York: Oxford University Press.

Ridley, Mark. 2004. *Evolution*. 3rd ed. Malden, MA: Blackwell Publishers.

Sober, Elliott. 1998. "Three Differences between Deliberation and Evolution." In *Modeling Rationality, Morality, and Evolution*, edited by Peter Danielson, 15. New York: Oxford University Press.

Spencer, Herbert. 1855. *The Principles of Psychology*. Williams and Norgate.

Sterelny, Kim, and Ben Fraser. 2017. "Evolution and Moral Realism." *The British Journal for the Philosophy of Science* 68 (4): 981–1006. https://doi.org/10.1093/bjps/axv060.

Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127 (1): 109–66. https://doi.org/10.1007/s11098-005-1726-6.

Walsh, Denis. 2000. "Chasing Shadows: Natural Selection and Adaptation." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of*

Biological and Biomedical Sciences 31 (1): 135–53. https://doi.org/10.1016/S1369-8486(99)00041-2.

———. 2015. *Organisms, Agency, and Evolution*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316402719.

———. 2021. “Environment as Abstraction.” *Biological Theory*, January. https://doi.org/10.1007/s13752-020-00367-2.

West-Eberhard, Mary Jane. 2003. *Developmental Plasticity and Evolution*. New York, NY: Oxford University Press.

Chapter 21

Adding Agency to Tinbergen's Four Questions

André Ariew[i], Karthik Panchanathan[ii]

Abstract

This year marks the 60th year anniversary of the publication of Niko Tinbergen's "On aims and methods of ethology" which remains influential among today's biologists and social scientists for its introduction of four questions for a complete explanation for animal behaviors. In this paper we argue that a large part of the lasting appeal to Tinbergen's four questions was (and still is) the methodological commitment to treating organisms as objects as opposed to purposive agents. It reinvigorated the discipline of ethology, allowing it to shed its teleological and anthropomorphic associations and to better cohere with a philosophy of science that favors inductive procedures, causal and mechanistic analytic techniques, and an emphasis on Darwinian explanations. While these features are still prized among today's biological social scientists, it ignores an important feature of many social organisms, that they are not merely objects, they are also purposive agents. We explore the implications that a shift from treating organisms as objects to treating them as agents has on both how we should interpret and answer Tinbergen's four questions. Updating Tinbergen's four questions with agency in mind only makes them more applicable to the biological investigation of animal behavior, but it also strengthens the value and applicability of biology-oriented research programs in the social sciences.

---

[i] Department of Philosophy, University of Missouri. Email: ariewa@missouri.edu

[ii] Department of Anthropology, University of Missouri. Email: panchanathank@missouri.edu

21.1 Tinbergen’s Four Questions

One of Darwin’s enduring legacies to the social sciences was to make legitimate the practice of biologizing human behavior. Humans are, after all, biological organisms and related to non-human animals by common ancestry. And so, social and behavioral scientists can learn a lot from ethologists. In the 1960s, Nobel laureate Niko Tinbergen (1963) categorized the biological study of animal behavior into four distinct kinds of questions, each with domain-specific goals and methodologies, which could and should nevertheless be integrated. Tinbergen’s four questions were:

- Causation: What causes the behavior?
- Survival value: What adaptive function does the behavior serve?
- Ontogeny: How is the behavior acquired?
- Evolution: How did the behavior become prevalent over evolutionary time?

Tinbergen formulated his taxonomy in reaction to what he understood at the time to be two kinds of dubious but common scientific practices: at one extreme, anthropomorphizing animal behavior and even ascribing purpose; and at the other, explaining animal behavior as mere reflexes to environmental cues. Tinbergen’s goal was to usher ethology into the scientific fold by emphasizing the role of objective observation and controlled experiment, the hallmarks—he argued—of good biology. In Tinbergen’s formulation, animal behaviors should be treated as organs, albeit complex organs. And just as biologists investigate adaptive organs through good scientific methods, so too should ethologists study animal behavior.

Sixty years later, behavioral and social scientists still celebrate the contribution that Tinbergen’s four questions made to the study of behavior, including human behavior. Nesse (2013) commemorates Tinbergen’s identification of ethology’s central questions as a “moment of discovery” for the biological sciences. Bateson and Laland (2013) honor Tinbergen’s legacy in promoting the four distinct areas of research and their integration through evolutionary theory. Kapheim (2019) applies Tinbergen’s framework to evaluate the current state of the study of eusocial insects, distinguishing those areas of investigation that have experienced rapid gains in knowledge and those in which we still know relatively little. In a recent special issue of *Philosophical Transactions B* (Legare and Nielsen 2020), a team of evolutionary social scientists report on their attempts to employ an integrative account of human ritual by using Tinbergen’s four questions of animal behavior. In the introductory article, Legare and Nielsen (2020) claim that collectively the work provided new avenues for theory and research into “this fundamental aspect of the human condition.”

Nevertheless, contemporary commentators agree that Tinbergen’s original questions and methods require modification and reinterpretation, and that the sciences need to do a better job of integrating and synthesizing the four questions. Bateson and Laland (2013), for example, declare that Tinbergen’s project of generating a comprehensive and integrated analysis over the four questions is far from complete in most areas (though they single out the science of bird song as a successful case).

One problem is that Tinbergen’s views are out of date with modern biological theorizing. To update Tinbergen’s approach requires more than the kinds of mere tinkering that contemporary advocates propose, especially if it is to be usefully applied to the investigation of human behavior. Tinbergen’s methodological prescription—to treat behaviors as organs, while useful for advancing ethology in the mid 20th century, is overly simplistic and reductionist, especially on its reliance on clear demarcations between genetic programs and environmental conditions. While his ideas were an improvement over the simple dichotomy of innate vs. learned or nature vs. nurture thinking of his mid 20th century interlocutors, he could not have known about the extended evolutionary synthesis (Oyama et al. 2001, Pigliucci and Miller 2010, Jablonka and Lamb 2014, Laland et al. 2015; see also Peterson 2017) and its emphasis on ontogenetic processes and their effects, including developmental constraints on evolutionary change, plasticity, epigenetics, and niche construction.

More broadly, Tinbergen's methodological prescription to view behaviors as organs is predicated on a view that animals are mere ***objects***: biological machines made of separable parts, passive and inert, structurally fixed, and acted upon by internal and external forces. While this perspective might be useful for many research questions, it misses a key feature of biological organisms. Organisms that exhibit behaviors are also ***agents***, not merely objects: still made of parts, of course, but self-organizing in their development and actively engaged in the modification of their environments. This shift from an object-oriented approach to an agency-oriented one has implications for how we interpret Tinbergen's four questions and how we should answer them. By updating Tinbergen's four questions with agency in mind we not only make it more applicable to the biological investigation of animal behavior, but we also strengthen the value and applicability of the biologizing the human mind research program because humans are paradigmatic agents. Critics of Tinbergen-inspired sociobiology and its subsequent fields were right that treating sentient beings (among other animals) as objects makes for an impoverished research program. But critics are wrong to think that biology cannot incorporate agency. (Likely these critics were misled by outdated biological theories like Tinbergen's.) That is not to say there are no limits to an agency centered version of Tinbergen's framework, but at least the debate can continue along fresh lines. For an instance of a potential limit, consider that the behaviors of some agents (at least humans) contain within them consciousness and subjective experience. An explanation of social behavior that does not account for these phenomena is, for some kinds of investigations, incomplete. An update that incorporates subjective experience may require an additional kind of question—perhaps Tinbergen's Fifth question. We will return to this question in the concluding section of this chapter.

Our overall motivation is programatic. We aim to refocus the debate about the legitimacy of the biologicizing minds research program beyond the usual questions of reductionism and determinism by showing how agency, along with its associated features of purposiveness, self-organization, and even consciousness could be incorporated. This paper initiates the project of addiing agency to Tinbergen's four with a focus on an exposition and criticism of Tinbergen's framework, including both Tinbergen's original formulation and on contemporary investigators—biologists and social scientists--who still espouse Tinbergen's object-oriented pre-suppositions. We then explore some of the ways in which an agency perspective adds to and changes how we think about Tinbergen's four questions.

21.2Tinbergen's Midcentury Ethology Program

Tinbergen's motivation for writing his 1963 paper was to evaluate the state of ethology, especially under the influence of his mentor and fellow Nobel laureate, Konrad Lorenz. Ethology was improving by attending to both good general scientific methodologies and insights from mid 20th century Darwinian biology. We should read Tinbergen's lessons in the context of his time. He devoted much of the essay criticizing animal behaviorists who veered from these practices. These included anthropomorphizers and teleologists who dragged down the legitimacy of the field by ascribing subjective experiences, intentions, or purposes as part of their explanation. According to Tinbergen, such entities were not legitimate objects of scientific study because they were not directly observable. At the other end of the spectrum, Tinbergen criticized reductionists of various stripes, including those steeped in conservative zoological traditions, who robbed the field of potential insights by over-emphasizing homology and anatomy while ignoring function, and behaviorists, who treated behaviors as simple reflexive reactions to external stimuli and failed to acknowledge the complexity of inner mechanisms and genetic programming. Tinbergen argued that Lorenz's ethology adhered to good scientific methodology by charting a middle path between these extremes.

Tinbergen's essay was and continues to be so influential because he managed to elevate ethology by articulating a philosophical view about what good scientific methodology entails along with a general account of what animal behavior is that made it suitable to scientific inquiry. To Tinbergen, ethology is a science that identifies behavioral patterns through inductive generalizations, facilitates causal analysis in answering all its relevant questions, and adopts both Darwin's theory of common descent and natural selection (we expound on these features below). To adopt these scientific principles, Tinbergen advocates an important auxiliary assumption about the ontology of behaviors: behaviors are organs no different than any other structural and physiological organ. In its fuller expression, animals possess species-specific adaptations which feature a complex ontogeny that involves an interaction of highly structured inner mechanisms (under genetic control) and external stimuli (in the sequence of environmental exposures during development). This is true regardless of whether these adaptations are behavioral, structural, or physiological in nature.

It is important to fully articulate these two aspects of Tinbergen's work—his philosophy of science and his account of animal behavior—because it provides the appropriate context to understand Tinbergen's four questions of animal behavior, why he chose them, and how he proposed ethologists generate scientific answers. This articulation is important for another reason—it shows how dated some of Tinbergen's views are. As Bateson and Laland (2013) note "almost every modern textbook on animal behaviour quotes his distinctions with approval" (p.1). If today's biologists and social scientists wish to adopt the same kind of rigor that Tinbergen demanded in the mid-20th century, then they should be willing to revise Tinbergen's philosophy of science and account of what constitutes animal behavior to reflect recent advances. Bateson and Laland (2013) argue, in their commemorative, that Tinbergen's scheme remains useful to this day as a heuristic but given developments in the sciences over the last 50+ years, the questions require a "more nuanced interpretation than is traditional" (p. 1). Tinbergen's schema doesn't need nuanced refinement—it needs an overhaul. To further progress in our understanding of animal (especially human) behavior, we need to admit that Tinbergen's account, looked at in the context of 21st century science was incomplete. It presupposed a view that individuals are mere objects, at the passive nexus of internal and external forces. Tinbergen did not consider organisms as agents that actively contribute to their conditions and generate behaviors according to their goals and needs. Agency is manifest in the entirety of the organic world and is most pronounced in the purposive behavior of humans. By adopting an agency view, we provide interpretations of Tinbergen's four questions that reflect not only a more complete biology but also a better biological underpinning for human social science. An agency view also reveals a limitation of Tinbergen's four questions—they cannot (by Tinbergen's own admission) apply to questions concerning subjective experience and consciousness, a goal for social scientists and a requirement in the humanities.

Nesse (2013) argues that emphasizing the controversies that Tinbergen's questions generate when applied to today's science "can obscure Tinbergen's accomplishment which remains vastly under appreciated." We agree that Tinbergen elevated ethology to a science by adopting good scientific methodological principles and practices and by endorsing Lorenz's radical step of regarding animal behaviors as organs in order to accommodate causal analysis. In the next section we will expound on this under-appreciated accomplishment. But, for Nesse, Tinbergen's other underappreciated accomplishment was to show that answers to all four questions are necessary for a complete biological explanation. We disagree. Without incorporating agency, we argue, Tinbergen's explanatory schema is incomplete. This is what we'll argue in the subsequent section of this paper.

21.3Tinbergen's Philosophy of Science

We have claimed that Tinbergen elevated the investigation of ethology by infusing it with good scientific methodology. And, he did so by adopting Lorenz's heuristic of treating animal behaviors as adapted organs. To unpack this let's begin with Tinbergen's prescription for good scientific methodology followed in the next section with an exposition of how adopting Lorenz's heuristic informed Tinbergen's formulation of his four questions. Later, we will show how treating behaviors as organs is, for better and for worse, part of the "objectancy" approach to ethology.

Induction: Ethologists practice inductive methods of data collection to support generalizations. To Tinbergen, the generalizations that mattered were the recognition that in the wild there exists an "enormous variety of animal behaviour repertoires" which were characteristic of individual species. Tinbergen's interlocutors missed the opportunity to ask questions like "why do these animals behave as they do" because they failed to even recognize their existence with their practice of singling out only a "handful of species which were kept in impoverished environments…and to proceed deductively by testing...theories experimentally."

Causal analysis: To Tinbergen, good scientists also adopt appropriate causal analytic techniques for answering each of the four questions. This is a pervasive theme in Tinbergen's essay. Adopting appropriate means of causal analysis allows biologists to dare to ask and even provide means to answer questions like "what causes this behavior?" and "what is this behavior good for?", while avoiding the looming specter of anthropomorphizing or teleology. Causal analysis takes on many forms in Tinbergen's analysis, including: mechanistic analysis of how a behavior contributes to a functional system, the careful investigation of cause-effect relations in trying to determine which of several effects promote survival value, a process of elimination to understand the differential effects of both the inner machinery and external environmental conditions in ontogeny, and the application of controlled selective pressures to determine the dynamics of evolution.

Finally, Tinbergen's ethology is thoroughly **Darwinian**, a requirement of any 20th century biology. Its scope and limits are co-extensive with Darwin's theories of common descent and natural selection. Natural selection provides the grounding for the "what for?" questions while Darwin's theory of common descent is at the heart of the elucidation of the course of evolution. Ethologists should judge the degree of evolutionary divergence by the degree of dissimilarity between current behaviors and their common ancestors. A rigid adherence to Darwinism is obviously one of the reasons why Tinbergen's four questions remain so attractive to today's biologists and evolutionary social scientists alike. But this rigid adherence to the Darwinism of Tinbergen's day also carries over the limitations that can lead to bad biology and bad social science.

21.4Tinbergen's View of Behaviors as Organs

Tinbergen formulated his four questions around a set of presumptions about animal behavior that allowed them to be the appropriate subject of good scientific theorizing. It is important to articulate the presumptions for the sake of understanding the motives and interpretations for each of his four questions. The most important is that animal behaviors are like organs. Earlier, we stated that Tinbergen's insight is that animals possess species-specific adaptations which feature a complex ontogeny that involves a complex interaction of highly structured inner mechanisms—under genetic control—and external stimuli, regardless of whether these adaptations are behavioral, structural, or physiological in nature. Let's now break this down into component parts to better appreciate how Tinbergen set the scope and limits of ethological investigation.

Tinbergen argued that behaviors are organs. This claim has two components. First, that behaviors are structural and physiological characteristics of animals as opposed to mental expressions. This is what makes ethology a science, as its investigation ranges over the physical features of objects, not the subjective experiences of agents. For Tinbergen, a good science ought to be based upon inductive methods where generalizations are supported from direct observations from both the field and from controlled experiments. Hence, ethologists should avoid ascriptions of subjective experiences and purposes ("teleology") to behaviors since both are, by their natures, not directly observable. Instead, ethologists should adopt the stance that animal behaviors are like organs which can be subject to inductive methods to uncover generalities, and causal analytic methods to generate explanations. Tinbergen emphasized the use of experiments to manipulate conditions and reveal important counterfactuals.

Second, like other organs, behaviors undergo ontogenetic development, a process that involves a complex interaction between an inner structure that is inherited from its parents and external features of the environment. Tinbergen stresses that there ought not be a methodological gap between ethology and neurophysiology as his interlocutors would have it. His interlocutors were simple behaviorists who thought behaviors as reflexes and hence over-emphasized the role of external stimuli. Tinbergen urged that behaviors are not reflexive expressions to external stimuli. Instead, they, like organs, undergo ontogenetic development.

Tinbergen also argued that behaviors, being organs, are species-specific adaptations. This claim also has two components. First, Tinbergen argued that behaviors are characteristics of species. That gives specificity to the scope of ethology, emphasizing categories of behaviors as opposed to individual expressions. Ethology is interested in behaviors that are characteristic of species, not the idiosyncrasies of individuals. It is part of an explanation for what makes, say, geese different than ducks, as opposed to what makes certain geese different from other geese. As Tinbergen put it, "each animal is endowed with a strictly limited, albeit hugely complex, behaviour machinery which (if stripped of variations due to differences in environment during ontogeny, and of immediate effects of fluctuating environment) is surprisingly constant throughout a species or population." (1969, p. 414) This argument has consequence for the science of ethology because it "positively facilitated causal analysis": "this awareness of the repeatability of behaviour has stimulated causal analysis of an ever-increasing number of properties discovered to be species-specific rather than endlessly variable." Tinbergen's identification of the phenomenon of interest as categories of behaviors aligns with his views about good inductive science appropriate for naturalists in the field. The descriptive task of ethology (the "return to nature") is to catalog the variety of species-specific behaviors so that they can be subjected to causal analysis and experimental manipulations for the sake of answering each of the four questions.

Second, by emphasizing that animal behaviors are adaptations, Tinbergen appeals to both (i) the current flourishing of animals and (ii) a causal explanation for their origins. This distinction is important for Tinbergen and the reason why he expressed one of the four questions in terms of "survival value" rather than "adaptation" as a means of interpreting the question "what is a behavior for?" (i) Behaviors aid their possessors to survive and reproduce in their natural surroundings: "It is through Lorenz's interest in survival value that he appealed so strongly to naturalists, to people who saw the whole animal in action in its natural surroundings, and who could not help seeing that every animal has to cope in numerous ways with a hostile, or at least uno-operative environment." (*Ibid*, p. 417) (ii) By referring to behaviors as adaptations, ethologists have a ready explanation in Darwin's theory of natural selection for their origin story that explains their prevalence among species. Most importantly, appeals to natural selection allow for scientifically minded ethologists to answer "what for" questions about behaviors without appeal to metaphysically suspect teleological forces. It also grounds the use of common descent to answer questions concerning the course of evolution current features underwent as a divergence from common ancestry.

21.5Tinbergen's Four Questions

Tinbergen's explanations for each of ethology's four questions presuppose his methodological commitments to what constitutes science and his 'adapted organs' account of animal behavior. Let's briefly go through each type of question, with an emphasis on how Tinbergen used his pre-suppositions to articulate how ethologists should provide scientific answers to each question.

1. The **causation** question is about "what causes the behavior?" To provide an appropriate scientific answer, one must avoid subjective, anthropomorphic, and teleological language. To say that "the animal attacks because it feels angry" is to ascribe a behavior that "can be observed by no one except the subject." Since we cannot observe an animal's feelings, the true source of the ascription must be derived from the human experimenter. Ethologists are often guilty of such teleological language. To refer to "innate reflexing mechanisms" is to characterize a mechanism in terms of achievement, making causal analysis difficult. Tinbergen prescribes treating a behavior like an organ that causally contributes and is causally integrated in sometimes very complex ways to a larger mechanistic context which provides its inputs and utilizes its causal outputs. Tinbergen envisions a future in which ethologists bridge the "no man's land" between ethology and neurophysiology through a hierarchy of causation, in which complex behaviors are broken down into component parts with the in-principle ability to continue the analysis down to molecular biology.

2. The **survival value** question allows us to distinguish from the various causal effects a behavior might have the one that explains "how the behavior works" by reference to its adaptive function. For example, a "releaser" is not merely "anything that provides stimuli" but "an organ characterized by a function." Darwinian natural selection is at the basis of questions of survival value because, like organs, species-specific behaviors owe their

prevalence to their adaptive function. Tinbergen notes that in post-Darwinian biology, questions about survival value got a bad reputation from the tendency of practitioners to make "uncritical guesses" from the "armchair", what Gould and Lewontin (1979) would later call "just-so stories". But, there are causal methods for testing function and survival value. Any hypothesis can undergo observational and experimental studies for the sake of revealing important counterfactuals. "Nest showing" among male sticklebacks can be shown to serve a causal function through the aid of dummies to control behavior and determine whether the behavior contributes to and is even indispensable for successful reproduction (p. 420). Tinbergen devotes a significant portion of this section on the distinction between past and current function. Past function explains how the behavior became prevalent, but the current function explains how an organism manages to survive in its current environmental state. Tinbergen argues from a methodological perspective that survival value for current environmental state should be established first since such hypotheses can be subject to observational and experimental studies. All together the hope is to provide a full story of cause-and-effect relationships to undergird the scientific explanation for what a behavior is for.

Bateson and Laland (2013) argue that Tinbergen's question should be understood today in terms of "current utility" rather than "adaptive significance" because it helps to emphasize the difference between a trait's etiological and current function. Nesse, instead, prefers "adaptive significance" over "current utility" because the latter invokes teleology of the noses are for supporting eyeglasses sort (2013, p. 682). Each side thinks that the dispute is more than terminological. We agree, but we side with Tinbergen who was: (i) articulate about the need to distinguish between etiological and current function (as Bateson and Laland urge), and (ii) was explicit about providing non-teleological answers. Tinbergen recognized the difference between the question "how did the species-specific evolve?" from the question "how do contemporary animals utilize their species-specific behaviors to flourish in their current environmental circumstances?" And, by promoting Darwinian evolutionary theory, ethologists can replace any teleological connotations with references to causal explanations about origins and current utility.

1. The **ontogeny** question investigates the "change of behavior machinery during development." Every aspect of Tinbergen's explanation for behavior development is infused with causal mechanistic analysis and a commitment to viewing behaviors as adapted organs. Explanations for how behavior develops involve first a distinction between the internal machinery and the external factors from the environment that make a difference, and second a method of "elimination" which involves varying environmental conditions to see if it makes a difference to the developing machinery. Labeling a feature as "innate" under this process is understood as a "negative" label, for it indicates that some number of external factors have been eliminated as candidates for making a difference in the development of the machine. For example, "if we raise male Sticklebacks in isolation from fellow members of its own species, subject them as adults to test with dummies, and find that they attack red dummies just as selectively as do normal males, we are entitled to say that exposure to red males cannot be responsible for the development of this selectiveness of response." (1969, p. 424) However, it does not follow that "innate" features do not require any "interaction with the environment". The appropriate conclusion is a description of the environmental aspects that were "shown **not** to be influential". It may be that certain environmental factors are required in other parts of the developmental process, or, possibly, it is required for proper functioning. For example, while juvenile Sticklebacks could be raised in darkness, they would not be fully functional. Innate in this context is the opposite of "environmentally-induced". Likewise, the interaction of internal machinery to environmental factors that are not eliminated from experimental manipulations are thought to serve to contribute to the internal machinery's developmental "programming". According to Tinbergen, there are two means by which organismal machinery is programmed in the individual: first, by evolutionary "trial-and-error-interaction with the environment which results in the specializations of the genetic instructions", and second, by "the ontogenetic interaction between the individual and its environment." Because programming could have its source in evolution, Tinbergen stresses that questions about causation of ontogeny are dependent upon the question of survival value, both rooted in Darwinian explanation.

2. **Evolution**. According to Tinbergen we should recognize that some behaviors are species specific and, like structures, can be studied comparatively between species, invoking Darwin's theories of common descent and natural selection. This is reasonable on the background assumptions that "individuals and populations differ as much in their hereditary behaviour 'blueprints' as in their hereditary structural blueprints"; and, "the genetic variation on which natural selection can act" is found in the hereditary blueprints.

The objective, then, for evolutionary explanation is to both elucidate the course of evolution and unravel its dynamics. The methods of the former are the same employed by the evolutionary taxonomist investigating physiological or structural characteristic. Beginning with a monophyletic group, the investigator judges the degree of evolutionary divergence by the degree of dissimilarity between innate traits ("of those characters that must be considered highly environmental-resistant ontogenetically"). Evolutionary dynamics are explained by both the methods of "geneticists" who identify the effects of mutations and cross-breeding on the evolution of the feature in question, and by the natural selectionist, who investigate either the survival value of the species-specific character or conduct controlled selection pressure experiments over a series of generations.

21.6Tinbergen's View of Organisms as Objects

It is important to put Tinbergen's program in historical context. Tinbergen's idea to regard behaviors as organs was a necessary step in the development of ethology as a scientific and, more importantly, a post-Darwinian discipline. Organs are objects with material constitutions, not subjective qualities as past animal behaviorists regarded behaviors. Hence, the ontological commitment to objects elevated ethology to a materialistic science, invoking the method of generalization of characters by inductive inferences over direct observations. It facilitated causal analysis by regarding the subject of study as part of the causal nexus of internal and external forces (as are ordinary objects or complex mechanistic ones), and allowed ethology to be subject to genetic analysis of ontogeny and evolutionary analysis of phylogeny. Darwinian evolution (especially post-Modern Synthesis with its emphasis on genetics) operates within the same ontological commitments. Denis Walsh labels this set of commitments "objectancy", for it treats individuals as material objects with intrinsic causal dispositions, or "propensities to behave in certain ways when they encounter certain external conditions." (2018, p. 3)

The objectancy approach harkens back to a Newtonian paradigm, that refers to the natural properties of objects and the external conditions that cause them to change. The natural state of an object is not to do much at all—in motion and at rest they remain in their initial states until subject to external forces. In fact, many of the relevant properties that explain an object's change exist independently of the object (p. 9). Consequently, there is a clear demarcation between objects, which largely remain unchanged, and the forces that cause them change or transformation (either internally or externally), which exist independent of the object. Tinbergen adopts the same distinction between organisms and the forces that determine development and evolution. He treats organisms as objects that remain unchanged unless they are subject to "influences" (Tinbergen's word) that exist independent of them. These influences might exist internally to them, as part of the "machinery" (again, Tinbergen's word, 1969, p. 424) that unfolds according to the complex interactions involving the species-specific genetically program, or the particular environmental conditions that the unfolding machinery encounters.

Further, Tinbergen's ethology relies on clear distinctions between organisms and the external conditions that determine their change. Organisms are demarcated by internal processes that generate variations of a genetical type which are then subjected to external forces, the environment, that selects among the variants. Both the internal processes and the external forces largely exist independent of the organism. The internal processes are dictated by a genetic program that is passed down to the organisms (from the outside) and was originally formed by external evolutionary processes. As Tinbergen said, the analysis of development is largely a matter of a process of elimination: vary environmental conditions and see if it makes any difference at all to the outcomes; if not, then, label the behavior as "innate". The internal-genetic explanation of development serves Darwinian evolutionary theory well because it explains differences between species—wolves, whales, and wallabies—in terms of genetic differences between them. Genetic differences are the stuff of evolution. The internal processes that genes control produce mutations and recombinations that are then tested for acceptability in the external environment (Lewontin 1985, p. 42).

An advantage of the object-approach to investigating organisms and their behaviors (as organs—another class of object) is that it provides us with a sense of regularity and order out of the chaos of individual variation. Newtonian physics is, again, the inspiration. Drop feathers from a height and they land in a scatter. But the scatter has a discernible pattern, beginning with a central cluster where most feathers land and radiating out where the fewer feathers lie. The Newtonian explanation distinguishes between regular and accidental causes. The regular causes are expressed as natural laws that determine the propensities of the object acting in the conditions of its state space. The center of the scattering is where each feather would land had it been subject to the main forces of gravitation, without interference from minor forces of wind and friction. The latter can be largely ignored, because the main objective is to see through the blooming, buzzing confusion of individual variation to find an underlying order. Tinbergen adopts the same approach for ethology. The unit of analysis is behavior that an individual expresses that is typical of its species. That allows Tinbergen to investigate "puzzling behaviour patterns" (1969, p. 412) in a systematic way, taking advantage of Darwin's theory of common descent. Behavioral patterns are treated as species-specific organs with features that are intricately adapted to their environment. By focusing on species-specific behaviors, ethologists can see past the buzzing, blooming confusion of individual variation and regard common regularities, the functions that adapted the feature to its environmental conditions.

Another way the objectancy stance provides order to the universe and its myriad of objects is by imposing a hierarchy of ascending functional systems where each system can be broken down into smaller sub-systems. The relation between the containing systems and the systems within them are a matter of causal connection—each sub-system produces an effect which together with its conspecifics produces the causal properties of a whole. Thinking about the universe in this way is advantageous to investigators of the natural world because of our natural cognitive ability to analyze and break down complex ideas into their simple parts recognizing how each part contributes to the whole. The critical assumption is that there are clear demarcations between objects from each other and from the containing system to which they contribute. On Tinbergen's view, animals are machines with internal parts and each part is seen as producing effects that contribute to a containing system. Individuals are a nexus of a variety of internal and external forces. Complexity has an easy measure by this machine style of analysis—to be more complex means that the system has more interlocking parts, sometimes with more feedback mechanisms. The point is, on Tinbergen's treatment, complexity is still a matter of cause/effect of the various components that make up the functional unit; there is no need to invoke teleological language to explain observed complexity. The view of individuals and their characters as objects facilitates the use of this venerable style of mechanistic analysis (hierarchies of systems within systems) that has served physics and chemistry well since at least the 17th century and is prominent in Tinbergen's questions.

In sum, Tinbergen's ethology, including his approach of providing answers to the four questions—causation, survival value, ontogeny, and evolution—is based upon treating organisms as objects with vestiges of a Newtonian paradigm: behaviors of interest are species-typical (i.e. typical of a type; variations within the type are accidental), organisms exist at the nexus of independent forces that determine both their internal development and external selection, explanations largely refer to what happens to organisms (rather than what organisms do), and how organisms and constituent parts contribute to the hierarchy of mechanistic systems.

21.7Organisms as Agents

The problem is that the objectancy approach to organisms and their features is an incomplete foundation from which to ground an investigation of life and behavior. The objectancy approach had the desired effect of facilitating causal analysis, but it gets a lot about ontogeny, causation, survival value, and evolution wrong. And it neglects important questions about behaviors generated by advances in developmental and evolutionary biology, as well as questions generated by thinking about the limits of the biologizing research program for human behaviors. Most importantly, objectancy ignores the role organisms and historical processes play in answering each of the four questions. Put another way, the objectancy approach, as Walsh (2018) puts it, ignores "agency". Agents are not mere objects. Objects remain the same until they are subject to forces. Agents have an additional feature from that of objects, they initiate their own changes. Tinbergen was so determined to avoid any association between ethology and teleology that he neglected to provide a means to explain patterns of purposive behaviors that are well-confirmed by good observational data. This entire object-oriented version of Darwinian biology, with its main goal of "facilitating causal analysis" is inadequate to the task of explaining animal behavior (both human and non-human) because it ignores what the agent's contribution is to causation, survival value, ontogeny, and evolution. As Lewontin put it: "classical Darwinism places the organism at the nexus of internal and external forces, each with its own laws, independent of each other and of the organisms that is their creation...The organism is merely the medium by which the external forces of the environment confront the internal forces that produce variation." (1985, p. 88 cited in Walsh 2018, p. 11)

21.8Adding Agency to Tinbergen's Four Questions

We have argued that Tinbergen's objectancy perspective is incomplete. In this section, we will explore some of the ways in which an agency perspective adds to and changes how we think about Tinbergen's four questions. This is not meant to be an exhaustive discussion, but instead a speculative introduction meant to motivate further exploration.

1. The **causation** question is about "what causes the behavior?" Recall that Tinbergen's goal was to put ethology on firm scientific grounds and to reject mystical appeals to anthropomorphism and teleology on the one hand and the overly reductionist approach of behaviorism on the other. Around the same time that Tinbergen was reformulating animal ethology, psychology was undergoing the cognitive revolution and abandoning its behaviorist past. In contemporary cognitive science, invoking concepts like intentions, goals, and desires are perfectly reasonable and perfectly scientific. Explanations at this level

can comfortably co-exist with explanations at other levels, including the neurophysiological. In fact, a complete psychological account should involve explanations at level of computation, algorithm, and implementation (Marr 1982). In Tinbergen's approach, the goal was to get as close to the level of implementation as possible; higher levels were considered less scientific. However, it's precisely these higher levels that feel more natural when talking about agency. At the level of computation, we can ask about the kinds of goals that agents have, or about the kinds of goals that components of their cognitive system have. Returning to our previous discussion of causation, from an agency perspective, it's perfectly scientific to say that "the animal attacks because it feels angry". This is not to deny any kind of lower-level, neurological understanding, but instead add to it.

2. The **survival value** question seeks to explain "how the behavior works" by reference to its adaptive function. In the object-oriented approach, the environment is supposed to present a population of organisms with some set of adaptive problems. The process of random mutation generates candidate solutions, in the form of variation in the population, and natural selection favors better solutions. Over time, the form of behaviors will be fashioned to adaptively function in the environment. While this textbook account no doubt captures many cases of adaptation, it is by no means the only way in which adaptation occurs. Take, for example, the process of "genetic assimilation" (Waddington 1953, West Eberhard 2003).

Let's imagine a mainland population of birds adapted to a generalist foraging strategy with a generalist's morphology to match. Suppose that a small group or even just a pregnant female are blown off course and end up on a faraway island. The ecology of this island does not match the mainland ecology to which the bird was adapted. In fact, let's imagine that the only edible foodstuff on the island is an orchid like plant with a long flowering body that provides nectar. At first, the birds will frantically search the island for edible items and find little success. Eventually, the birds will learn about that these flowers and extract nectar from them. Assuming there is no social learning in this species, each generation of birds must learn to feed on the nectar of these flowers. This process of learning within each generation sets up a recurrent phenotype-environment match. But this match is entirely driven by the goal-directed actions of the birds (i.e. seeking nutritious foods from the environment). Now, imagine there is genetic variation in this population, as there must be. Any mutation that changes beak morphology to better extract nectar from these long flowers will be favored by natural selection. Likewise, any changes to gut morphology to extract calories more efficiently from nectar will be favored. And, natural selection will also favor learning systems that are prepared to associate those particular flowers with food. Over time, the bird population will become behaviorally and morphologically adapted to being a specialist feeder on this flower. However, the process by which this happened involved the purposive and goal-directed behavior of birds, generation after generation. These birds were not passive objects that were transformed by the processes of mutation and natural selection. Instead, these birds created the conditions that led mutation and natural selection to reconfigure their behavior and morphology.

3. The **ontogeny** question investigates the "change of behavior machinery during development." Tinbergen's view of ontogeny has all the hallmarks of a commitment to viewing organisms as objects. It begins with the genetic program sourced from the outside—the parental organisms. The process of development is largely a matter of mapping how this species-typical genetical program combines with the set of environmental conditions to which it is exposed. On this view, the organism is passive; development happens to it. As Lewontin (2001) argues, this approach ignores the myriad of ways in which organisms play an active role in determining how the "environment" influences their development.

    - Organisms **determine** which elements of the external world are put together to make up their environments. A Phoebe and a thrush can both co-exist in a plot of land, but it doesn't mean they share the same environment. A Phoebe uses grass for nesting, has no use for the stones that the thrush uses as an anvil.

- Organisms actively **construct** a world around themselves. Earthworms make burrows in land that are filled with the aqueous substance similar to that of an ocean from which their ancestors thrived for 50 million years (Gilbert and Ebel, 2015, p. 466).
- Organisms **alter** and transform matter and energy, passing along one form to others which then can be used as a resource (2001, p. 55). Mycorrhizae is a symbiotic relationship between plants and fungi which have effects on individual life cycles, gene expression and inter-species systems of energy transfer. The fungus benefits from direct access to essential carbohydrates that the root tissue provides. In exchange, orchids acquire carbon that the fungi provide, without which the seeds could not germinate (Gilbert and Ebel, 2015, p. 86). The wood wide web refers to the energy network formed by the mycelia of fungi that colonize a roots of various plant species. The result of this complex and reciprocal transformation of energy is distribution system and even a communication network. This is a remarkable example, because out of features of two different kinds of agents, plants and fungi, is a third order agency constructed out of the symbiotic relationship. There is fledging research program around the idea that organisms are really "holobionts", composed of an ecosystem involving a variety of systems sharing and outsourcing some essential functions.
- Organisms **modulate** the statistical properties of external conditions. Plants photosynthesize when energy is available during the day but not the night. Desert plants may have an opportunity to germinate and grow only on one out of five years. Modulation is an individual's way of flourishing despite the fluctuations in availability of essential resources. Rituals like feasts and potlatches, artifacts like grain sheds, freezers, and even the creation of currency are important human manifestations for the need to modulate environmental fluctuations.
- Organisms **transduce** one kind of physical signal to another one. Organisms do not simply receive information from the signals they encounter from the world but they convert the signal into a different kind so it can be perceived by the organism's functioning system. Mammals convert rise in air temperature by the hypothalamus to an endocrine signal which causes changes in a number of chemical, neural, and

anatomical activities. Ironically, Lorenz's and Tinbergen's work provided breakthroughs in understanding some of these kinds of signal transductions, but Tinbergen did not recognize the theoretical ramifications against the objectancy approach: organisms are not passively responding to external conditions, but actively commingling with their environment, adjusting in ways that enhance their flourishing.

4. The question of **evolution** entails the unraveling of the evolutionary dynamics that led to the current behavioral form. In the objectancy perspective, and as with the question of survival value, this amounts to treating the population of organisms like a bunch of billiard balls subject to various evolutionary forces. In the agency approach, organisms become active participants in the processes that shape their evolutionary histories. As discussed in the section on ontogeny, organisms act upon the environment just as the environment acts upon them, making apportioning causal responsibility much more interactive and holistic. As Walsh (2015, p. 157) puts it, "Just as the actives of the system as a whole are the causal consequences of the activities of the component parts, so too the activities of the component parts are controlled and regulated by the system as a whole." When the aggregate actions of a population of organisms result in measurable changes to the environment, the adaptive landscape has been altered. In this way, organisms shape the environments in ways that result in novel selection pressures acting on subsequent generations, a process called "niche construction" (Odling-Smee et al. 2003). A canonical example of this process is provided by beavers constructing dams across river systems, thereby creating lakes, and changing the flow of water through the environment. These changes not only affect the subsequent evolution of beavers, but also of other organisms in the environment.

21.9 Objects Languish, Agents Flourish

Recall that Tinbergen formulated his object-approach to organisms and behaviors in part to combat mystical teleological thinking inherent in the ethology literature at the time. However, as Okasha (2018) points out, there are good reasons to treat organisms as agents, regardless of their cognitive abilities: (i) organisms are the locus of goal-directed activities, (ii) organisms exhibit "behavioral flexibility", (iii) organisms possess adaptations that "appear designed for a purpose". Lewontin's (2001) description of the various things that organisms actively do in their environment provides many instances of what Okasha is talking about. Examples of goal directed and flexible behaviors including determining which elements of the external world are put together to make up "their" environment, and actively constructing a world around them. Okasha adds examples of courtship behavior, way-finding or homing, and food storage and retrieval, which are commonplace in nature.

To be sure, nothing in Okasha's three reasons for adopting agency necessitates an overhaul in Tinbergen's approach to answering the four questions. In fact, Okasha means to demonstrate that all three rationales are defensible from conservative biological practices. However, as Walsh (2015) points, treating organisms as purposive, self-regulating, goal-directed entities turns traditional Darwinian thinking on its head: "there is no need to think of selection as a discrete cause that introduces adaptive bias into population change." (p. 157) That is to say, a consequence of taking the agency view seriously is that Darwinian evolution is no longer the theoretical structure at the center of explaining adaptive change, the developmental system is.

To see how deeply this upsets Tinbergen's objectancy approach to answering his questions recall how Tinbergen treats the question of survival value and evolution of adaptive behaviors. On the evolutionary approach, behaviors are adaptive because they confer fitness-enhancing benefits. To answer evolutionary questions about a feature's origins, we invoke Darwinian theory: adaptations are genetically inherited variants that in the past conferred fitness-enhancing benefits. Over time the direction of evolutionary population change favored these variants. To answer questions about current survival value, Tinbergen stressed that we need to recognize that the current selective regime need not be the same as what a population experienced in the past. That's why Tinbergen introduced survival value as a distinct question from evolutionary history. But the underlying mechanism is the same—adaptation explained by Darwinian selection. But, by putting agency at the center of the investigation, you detach adaptation from its genetical, fitness-enhancing interpretation, and replace it with a broader notion of “flourishing”. Flourishing in this sense is not a throw-back to Tinbergen’s spiritual interlocutors. Rather, the concept of an agent’s flourishing is grounded in modern-day views about ontogeny (as we have described, above). As Walsh (2015) puts it: “In development organisms orchestrate, integrate, accommodate and negotiate the various causal influences from genes, genomes, epigenetic factors, cells, tissues and environments in the production of a stable, highly adaptive responsive entity. That, in turn, requires acknowledging the significance of organismal purposiveness for evolution.”

21.10 Bridging Evolution and the Social Sciences with Agency

One advantage of this theoretical freeing of flourishing from its evolutionary (fitness-enhancing) interpretation is it makes the program of biologizing behavior more palatable for traditional social sciences. In addition to criticizing Tinbergen’s approach to the four questions as an insufficient biology, we also criticized it an insufficient social science. To see why, let’s start with a critical look at those precincts of the social sciences that have adopted evolutionary approaches and in implicit or explicit ways engage in human ethology within Tinbergen’s framework.

The first iteration of evolutionary social science was in the form of sociobiology (Wilson 1975). This paradigm was rightly criticized for reductionism gone too far in its attempt to explain every instance of human behavior in terms of fitness maximization (Gould and Lewontin 1979). This was a level of reductionism that might have made even Tinbergen cringe. In this approach, only one question matters: How does the behavior increase survival value? In this framework, there is no scope for mechanisms as there is apparently a direct causal connection between adaptive problem and fitness-enhancing behavior. Likewise, development and evolution drop away. In a way, the sociobiology approach adopts an agency perspective, but a strange kind of agency in which organisms, including humans, seek to maximize their inclusive fitness, ultimately serving their genetic masters.

In the wake of human sociobiology arose “three styles” of doing evolutionary social science (Smith 2000): human behavioral ecology, evolutionary psychology, and cultural evolution. While these disciplines represent much more sophisticated approaches to the study of human behavioral and social science, they are still firmly rooted in what Walsh calls the objectancy perspective. In their own ways, each of these disciplines ignore the role for agency in human affairs.

1. Human behavioral ecology (Borgerhoff Mulder 1991, Cronk 1991, Smith and Winterhalder 1992) drew inspiration from economics and posited the ability for humans to behave in ways that were optimized for their environment in terms of fitness maximization. This approach typically invokes the “phenotypic gambit” (Grafen 1984) and “black boxes” the mechanisms underlying behavior, including its acquisition and evolution. As with sociobiology, there’s a kind of agency here if we’re willing to assume that organisms are trying to maximize fitness. In this case, agency doesn’t reside within organisms; instead, the agents seem to be the underlying genetic programs which seek to maximize their own fitness by having their host organisms optimize behavior in ways that correlate with fitness maximization. However, most practitioners of human behavioral ecology do not make this assumption. Instead, the phenotypic gambit is taken as an epistemological approach, not an ontological commitment. As such, the approach has little to say about the issue of agency.

2. In evolutionary psychology (Barkow et al. 1992, Pinker 2003, Gaulin and McBurney 2003, Buss 2014), there’s no agency left. Instead, all causal force is attributed to natural selection which shapes the cognitive and behavioral mechanisms to behave in adaptive ways. In this view, development is similar to Tinbergen’s sense of development, a species-typical genotype is exposed to a set of environmental conditions which results in an unfolding process of development.

3. Cultural evolution (Cavalli-Sforza and Feldman 1981, Boyd and Richerson 1985) draws inspiration from the “blank slate” view of human nature. In this view, natural selection

shaped the capacity for cultural transmission, thereby creating the conditions for a second evolutionary process that affects human evolution: culture. If in evolutionary psychology it was natural selection that adapts humans to their environment, in the cultural evolution approach, it's culture. There are various forces of cultural evolution which, over time, adapt a population of humans to their environmental conditions. This approach tends to treat individual humans are passive parts of this process, blank slates upon which culture can inscribe norms, values, and behaviors.

While these various schools of evolutionary social science have been successful at guiding the study of human behavior, they seem to leave little room for agency, at least agency within individual organisms. This matters because humans are the most complex types of agents out there. If an agency approach to ethology results in better biology, then it seems to be a requirement for any attempt at an evolutionary social science. Furthermore, interest in agency has been an important part of the social sciences, especially in the last fifty years. If the goal of the evolutionary social sciences is further penetration into the social sciences and humanities, then it seems to be of paramount importance to offer an evolutionary approach to human ethology with agency at its heart.

Obviously, exploring the ways in which evolution and agency can work together in fashioning a new kind of social science is a daunting task. Here, we focus on one example to see how an agency approach may help to narrow the gap between evolution and the social sciences. We'll consider ritual as this has long been a topic of interest in the social sciences, and a recent issue of *Philosophical Transactions* (Legare and Nielsen 2020) has focused on how Tinbergen's four question approach can contribute to the study of ritual.

Rituals are a series of actions, which are regularly repeated over the years and generations by a community of individuals, and which embody the beliefs of that group of people and foster a sense of community. The study of ritual has a long and deep history in many social science disciplines, especially sociology and anthropology. For example, Durkheim and later functionalist anthropologists of the mid 20th century focused on the socially integrative functions of rituals. For anthropologists like Clifford Geertz and Victor Turner, rituals were important in terms of their symbolic meaning to practitioners.

In the recent special issue on ritual, the authors apply Tinbergen's four questions to the study of ritual. The claim is that this approach will revolutionize the study of ritual. However, as with the evolutionary social science disciplines discussed above, the authors in this special issue adopt an objectancy perspective. The humans engaged in these ritual activities are relatively passive participants. A well-known example of this approach involves the work of Richard Sosis. As Sosis and Bressler (2003) point out, many collective rituals involve costly displays on the part of the practitioners. Drawing on costly signaling theory, they argue that one of the main functions of these costly rituals is to selectively filter out those individuals who are not committed to the long-term goals of the community. Many communities are sources of cooperation. The problem with cooperation is the presence of free riders, those who partake in the gains of cooperation without contributing to it. The authors argue that costly ritual displays act as a filtering device. Those who are willing to pay the costs of cooperation are also willing to bear the costs of the ritual. Those who seek to free ride on the hard work of others are less willing to incur the costs of rituals. We believe that there is much to this argument. However, this approach treats individuals as coming in one of two fixed types: cooperators and free riders. Some process of cultural evolution, external to the cultural practitioners, has created the institution of costly rituals as a way of filtering among individuals, admitting cooperators into the community and rejecting free riders.

But this is not the only function of ritual. Malinowski, an early figure in anthropology, argued that rituals give humans a comforting sense of control, especially during times of uncertainty. In this view, humans have beliefs about how the world works and engage in ritual behavior in order bring about useful interventions. The work of Evans-Pritchard (1937) is instructive. Evans-Pritchard argued that the Azande had two kinds of explanations for unfortunate events, one materialistic and one intentional. For example, suppose that a child suddenly falls out of a tree and dies as a result. The Azande would certainly agree that the death resulted from the fall. However, they would ask another question: “Why was it this boy that fell from the tree and not some other boy?” This second question involves a different kind of answer, one often involving the practice of witchcraft. Someone in the village must have wished ill for that boy and, as a result, the boy falls from the tree. While we may not agree with this causal logic, it has real world consequences for the Azande, including rituals to uncover who the witch was. Explaining this kind of ritual is very different than the kind of explanation Sosis and Bressler offer from costly rituals. An important aspect of rituals is to make sense of the world and gain some sense of control over it. This is not the kind of thing that an object would do. This is the kind of thing an agent does.

This approach to studying behavior is to take the program of biologizing behavior in a different direction than what Tinbergen envisioned. Rather than regarding behaviors as parts of objects for the sake of applying a mechanistic methodology, regard them as expressions of their goal-directed, behaviorally flexible, purposive activities. This is closer to what human social sciences endeavor to do.

21.11 Tinbergen’s Fifth Question

While an agency-focused approach to Tinbergen’s four questions would do a lot to make evolutionary approaches more palatable to some social scientists, it may not be enough for others. One aspect of agents, as opposed to objects, is their ability to actively participate in their own transformational processes and in modifying their environments. We can think about this as one kind of agency. But there’s another kind of agency, especially for humans: consciousness. Conscious agents not only act upon the world, but they have a subjective experience of themselves, their world, and what they are doing to the world and what the world does to them. Mary the color scientist not only sees the wavelengths of light corresponding to the color red, but upon seeing an apple for the first time she has the subjective experience of red (Jackson 1982). Nagel’s (1974) question “What is it like to be a bat?” seems to offer a difficult challenge for the Tinbergen approach to studying behavior. While we can map the mechanisms, ontogeny, function, and evolution of echolocation, we will never be able to experience what that form of navigation is like. The degree to which this kind of consciousness or subjective experience makes a difference in explaining and predicting how organisms, especially humans, behave may make a difference. But it's not clear how to deal with this kind of phenomenon within Tinbergen's framework. Perhaps this requires another kind of question: Tinbergen's Fifth.

References

Barkow, Jerome H., Leda Cosmides, and John Tooby, eds. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press, 1992.

Bateson, Patrick, and Kevin N. Laland. "Tinbergen's four questions: an appreciation and an update." *Trends in ecology & evolution* 28.12 (2013): 712–718.

Borgerhoff Mulder, M. "Human behavioural ecology." In *Behavioural Ecology: An Evolutionary Approach*, edited by J. R. Krebs and N. B. Davies. Blackwell, 1991: 69–98.

Boyd, Robert, and Peter J. Richerson. *Culture and the Evolutionary Process*. University of Chicago Press, 1985.

Buss, David M. *Evolutionary Psychology: The New Science of the Mind*. Psychology Press, 2014.

Cavalli-Sforza, Luigi Luca, and Marcus W. Feldman. *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press, 1981.

Cronk, Lee. "Human behavioral ecology." *Annual Review of Anthropology* 20 (1991): 25–53.

Evans-Pritchard, Edward E. *Witchcraft, oracles and magic among the Azande*. Vol. 12. London: Oxford, 1937.

Gaulin, S.J.C. and D.H. McBurney. *Evolutionary Psychology*. Pearson, 2003.

Gilbert, Scott F., and David Epel. *Ecological Developmental Biology: The Environmental Regulation of Development, Health, and Evolution.* 2nd edition. Sunderland, Massachusetts, U.S.A, 2015.

Gould, Stephen Jay, and Richard C. Lewontin. "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme." *Proceedings of the royal society of London. Series B. Biological Sciences* 205.1161 (1979): 581–598.

Grafen, A. "Natural selection, kin selection and group selection." In *Behavioural Ecology: An Evolutionary Approach (2nd Edition)*, edited by J.R. Krebs and N.B. Davies. Sinauer Associates, 1984. 62—84.

Jablonka, Eva, and Marion J. Lamb. *Evolution in Four Dimensions, Revised Edition: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. MIT press, 2014.

Jackson, Frank. "Epiphenomenal qualia." *The Philosophical Quarterly* 32.127 (1982): 127-136.

Kapheim, Karen M. "Synthesis of Tinbergen's four questions and the future of sociogenomics." *Behavioral ecology and sociobiology* 73.1 (2019): 1–15.

Laland, Kevin N., et al. "The extended evolutionary synthesis: its structure, assumptions and predictions." *Proceedings of the Royal Society B: Biological Sciences* 282.1813 (2015): 20151019.

Legare, Cristine H., and Mark Nielsen. "Ritual explained: interdisciplinary answers to Tinbergen's four questions." *Philosophical Transactions of the Royal Society B* 375.1805 (2020): 20190419.

Lewontin, Richard C. *The triple helix: Gene, organism, and environment*. Harvard University Press, 2001.

Lewontin, R. C. "The Organism as Subject and Object of Evolution." In R. Levins and R. Lewontin, *The Dialectical Biologist*. Cambridge, MA: Harvard University Press. 1985: 85-106.

Marr, D. (1982), *Vision: A Computational Approach*, San Francisco, Freeman & Co.

Nagel, Thomas. “What is it like to be a bat.” *Readings in philosophy of psychology* 1 (1974): 159-168.

Nesse, Randolph M. “Tinbergen’s four questions, organized: A response to Bateson and Laland.” *Trends in Ecology & Evolution* 28.12 (2013): 681–82.

Odling-Smee, F. John, Kevin N. Laland, and Marcus W. Feldman. *Niche Construction: The Neglected Process in Evolution*. Princeton University Press, 2003.

Okasha, Samir. *Agents and goals in evolution*. Oxford University Press, 2018.

Oyama, Susan, Paul E. Griffiths, and Russell D. Gray, eds. *Cycles of Contingency: Developmental Systems and Evolution*. MIT Press, 2001.

Peterson, Erik L. *The Life Organic: The Theoretical Biology Club and the Roots of Epigenetics*. University of Pittsburgh Press, 2017.

Pigliucci & Muller. *Evolution, The Extended Synthesis*. MIT Press, 2010.

Pinker, S. *The Blank Slate: The Modern Denial of Human Nature*. Penguin Books, 2003.

Smith, Eric Alden. “Three styles in the evolutionary analysis of human behavior.” *Adaptation and Human Behavior: An Anthropological Perspective* (2000): 27—46.

Smith, Eric A., and Bruce Winterhalder, editors. *Evolutionary Ecology and Human Behavior*. Routledge, 1992.

Sosis, Richard, and Eric R. Bressler. “Cooperation and commune longevity: A test of the costly signaling theory of religion.” *Cross-cultural research* 37.2 (2003): 211-239.

Tinbergen, Niko. “On aims and methods of ethology.” *Zeitschrift für tierpsychologie* 20.4 (1963): 410–433.

Waddington, Conrad H. “Genetic assimilation of an acquired character.” *Evolution* (1953): 118–126.

Walsh, Denis M. *Organisms, agency, and evolution*. Cambridge University Press, 2015.

West-Eberhard, Mary Jane. *Developmental plasticity and evolution*. Oxford University Press, 2003.

Chapter 22

Cultural evolution needs human behavioural ecology and cultural transmission approaches

Alberto J. C. Micheletti[i†], Eva Brandl[ii†], Hanzhi Zhang[iii], Sarah Peacey[iv] & Ruth Mace[v*]

Abstract

Human culture changes over time and varies across space. Two main approaches to study cultural evolution have developed in the last fifty years: human behavioural ecology and a suite of perspectives centred on the role of cultural transmission. The latter are often confusingly referred to with the name of the phenomenon they are trying to explain, ‘cultural evolution’. We argue that this is unhelpful and is generating confusion, including the claim that human behavioural ecology disregards cultural evolution. The aim of behavioural ecology is to explain human behaviours, and the vast majority of them are at least to some extent cultural. In addition, culture forms part of the ecology that determines the costs and benefits associated with adopting a behaviour. Thus, human behavioural ecologists have studied cultural evolution from the very beginning, even though they have not focussed on social learning. We explore three examples in detail: kinship systems, religious institutions, and witchcraft belief. We then use the framework offered by Tinbergen’s [1963, Z Tierpsychol, 20(4), 410-433] four evolutionary questions about behaviour to explain how human behavioural ecology and cultural transmission approaches can fruitfully coexist and complement each other. Moreover, we discuss several difficulties with cultural transmission approaches and highlight how the human behavioural ecological view of cultural evolution sometimes diverges from them. We conclude by suggesting that the field can move forward and achieve greater synthesis by exploring how

selective processes acting on biological fitness differ from those acting on cultural fitness – and how the two might interact in the cultural evolution of human behaviours.

---

[i] Department of Anthropology, University College London, 14 Taviton St, London WC1H 0BW, United Kingdom.

[ii] Department of Anthropology, University College London, 14 Taviton St, London WC1H 0BW, United Kingdom; BirthRites Independent Research Group, Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany.

[iii] Department of Anthropology, University College London, 14 Taviton St, London WC1H 0BW, United Kingdom.

[iv] Department of Anthropology, University College London, 14 Taviton St, London WC1H 0BW, United Kingdom.

[v] Department of Anthropology, University College London, 14 Taviton St, London WC1H 0BW, United Kingdom.

† First authors, contributed equally

* Corresponding author

22.1 Introduction

Human culture changes over time; it evolves. A considerable body of work has developed suggesting that it does so in ways that are at least partially analogous to organic evolution, notwithstanding the notorious problems with defining culture (Driscoll 2017; Lewens 2012, 2020). Cultural variants – socially-learnt behaviours, including the production of artifacts (Birch 2017) – are transmitted between individuals (inheritance with modifications), vary within populations, and between populations if they become widely shared cultural norms (variation), and in some cases result in differential survival for their carriers (selection) (Boyd and Richerson 1985; Cavalli-Sforza and Feldman 1981; Creanza et al. 2017; Mesoudi 2011). Darwin already recognized this in relation to language change over time, observing that language evolution is phylogenetic and that words are more likely to survive if they are easy to remember (Darwin 1877, p. 113).

The application of evolutionary thinking to cultural change took a long time to reach scientific maturity. The concept of cultural evolution was first elaborated by social anthropologists in the 19th century, who proposed grand narratives of history that assumed that humanity passed through progressive stages of cultural development, from savagery through barbarism on to civilization (Carneiro 2003). These perspectives were not as naively unilinear as is often claimed and did not deny the possibility of 'degradation' or 'regression' (Carneiro 2003), but they nevertheless assumed that, by and large, societies go through successive stages of development, which they do not (Currie et al. 2010). This differs from the tree-like view of evolution in Darwin's (1859) theory and these early cultural evolutionists never proposed a formal theory of variation, inheritance, and selection. At the same time, early social anthropologists defined culture as "that complex whole which includes knowledge, belief, art, morals, law, custom, and any other capabilities and habits acquired by man as a member of society" (Tylor 1903). In other words, 'culture' comprised just about everything that humans do but was also conceptualized as being exclusive to the human species, which further hindered a truly scientific investigation of culture. Most social and cultural anthropologists later abandoned the concept of cultural evolution as too reductionist and have debated numerous competing definitions of 'culture' over the years (Kuper 1999). Many of these incorporate the notion that cultures are made up of "patterns of values, ideas, and other symbolic-meaningful systems" (Kuper 1999, p. 69) (citing Kroeber and Parsons 1958) and deliberately disregard evolutionary considerations, with some exceptions (see Carneiro 2003).

Researchers only began to explore the evolution of culture with scientific rigour in the 1970s and 1980s, when there was an explosion of interest in the application of evolutionary theory to human behaviour (reviewed in Laland and Brown 2011). This was initially triggered by the publication of E. O. Wilson's (1975) *Sociobiology*. Some of the ideas expounded by Wilson proved to be controversial and, for this reason, the term sociobiology was soon abandoned. In response to this debate, three evolutionary approaches to the study of human behaviour emerged (Laland and Brown 2011). The first is centred on cultural transmission through social learning: it comprises a suite of perspectives and associated theories that, over time, have become known collectively with the name of the phenomenon they are trying to explain, "cultural evolution" (Boyd and Richerson 1985; Cavalli-Sforza and Feldman 1981; Mesoudi 2011). The second, human behavioural ecology, focuses on biological fitness and whether a given behaviour is adaptative in its ecological context (Nettle et al. 2013). The third, evolutionary psychology, focuses on universal cognitive adaptations supposed to have a genetic basis (Tooby and Cosmides 1990). Its proponents tend not to investigate why culture varies between populations, focusing instead on species-specific cognitive mechanisms shared by all members of our species. For this reason, we do not discuss this perspective here. Also, in the context of "cultural evolution", we will not discuss memetics (Dawkins 1976), as this theory has never generated a rigorous research programme (see Laland and Brown 2011).

22.2The "cultural evolution" approach focuses on transmission

"Cultural evolution" researchers suggest that culture constitutes a second inheritance system in addition to genes (Boyd and Richerson 1985; Cavalli-Sforza and Feldman 1981). While many behaviours can be influenced potentially both by genetically transmitted and culturally transmitted factors (e.g. dairy farming), most human behaviours are believed to be heavily influenced by culture. Boyd and Richerson (1985, p. 5) define culture as information acquired "through teaching, imitation, and other forms of social learning". Social learning refers to what we learn from others, as opposed to individual learning, that is information that individuals acquire on their own, interacting with their environment. Social learning can be adaptive, meaning that it generates fitness benefits for the learner, because it reduces the cost of individual learning. Instead of having to painstakingly 'invent' or 'discover' all the relevant information about the environment and the attendant survival skills by themselves, social learners can rely on what others are already doing (Boyd and Richerson 1985). This is particularly beneficial when the environment is stable, when the behavioural solutions that others have come up with are still relevant, and when solutions are complex and the cumulative knowledge of previous generations can be built upon. Socially learned behaviours can then provide individuals with solutions to commonly encountered survival problems and buffer them against natural selection (Boyd and Richerson 1985), although not all socially learnt traits are necessarily fitness relevant.

Treating culture as an information system has allowed "cultural evolution" researchers to draw parallels between cultural and organic evolution, and to adapt some of the modelling tools of the latter to the former (Boyd and Richerson 1985; Cavalli-Sforza and Feldman 1981; Mesoudi 2011). Due to differences between cultural and genetic inheritance, culturally transmitted traits evolve differently from genetically controlled ones. While genes only spread vertically (parent to offspring), social information may also spread obliquely (older generation to younger generation, excluding parents) and horizontally (peer to peer, within the same generation) (Cavalli-Sforza and Feldman 1981). The transmission process can also take different forms. These include: local enhancement, where a learner's attention is drawn to an object used by another individual and then the learner interacts with the relevant object themselves; observation, where a learner observes another individual perform an activity; emulation, where a learner observes another individual perform some activity and then performs actions that achieve a similar effect; and imitation, where a learner observes another performing an action and then copies the entire action sequence step-by-step (Hoppitt and Laland 2013).

Some "cultural evolution" researchers argue that these transmission processes can be biased in different ways, for example in favour of copying the majority in a social group (conformity bias) or copying particularly prestigious individuals (prestige bias) (Boyd and Richerson 1985; Kendal et al. 2018; Laland 2004). In their view, cultural change is heavily shaped by these transmission biases and thus they are deemed crucial to our understanding of this process (Kendal et al. 2018). They suggest that these biases, especially conformist learning, can stabilise behaviours within groups and allow the spread of more successful norms through a process of cultural group selection (Boyd and Richerson 1985; Henrich 2004; Richerson et al. 2016). In recent years, a heated debate has raged over whether cultural group selection can be a significant driver of cultural change (see D. Smith 2020 for a detailed review).

Those working on cultural transmission approaches did not make much of the distinction between micro- and macro-evolution. And thus they did not initially engage with cultural phylogenetic approaches (Mace and Pagel 1994), other than to be somewhat hostile to the idea (Boyd et al. 1997). This was largely based on the idea that they felt cultural phylogenetics was invalidated by horizontal transmission. However, horizontal transmission within cultures does not invalidate a phylogenetic approach – indeed it strengthens it. Horizontal transmission between cultures somewhat presumes the existence of a 'cultural tree' in the first place, or the term is meaningless. Cultural comparison has always been fundamental to anthropology, and the use of formal phylogenetic comparative methods started to be used by some human behavioural ecologists and evolutionary anthropologists to study cultural evolution from the mid 1990s on.

22.3The human behavioural ecology approach focuses on biological fitness

Behavioural ecology is the study of animal behaviour from an evolutionary perspective, and it has been an exceptionally successful research framework for behavioural studies over the past 50 years (Davies et al. 2012; Krebs and Davies 1978). Human behavioural ecologists aim to apply the principles of behavioural ecology to humans (Borgerhoff Mulder 1991; Borgerhoff Mulder and Schacht 2012; Mace 2014; Nettle et al. 2013; E. A. Smith and Winterhalder 1992; Winterhalder and Smith 2000). Their main goal is to explain the diversity of human behaviours in terms of their biological function, that is whether and how they are adaptive in the context of the local environment. Notice that this does not mean that behavioural ecologists maintain that all behaviours are adaptive: some might be maladaptive (see Section 5.4), and others might be neutral with respect to fitness (see Section 8).

Evolutionary biology has shown that natural selection leads to individuals that appear designed as if to maximise not the fitness of their group, but their own inclusive fitness, that is the sum of an individual's reproductive output and the reproductive output of their relatives, weighted by the level of relatedness between them (Darwin 1859; Fisher 1930; Grafen 2007, 2014; Hamilton 1964). Behaviours are therefore expected to be shaped by the individual's inclusive fitness interests. Specifically, a behavioural strategy can evolve and be stable if the marginal cost of adopting it for the actor (c) is outweighed by the marginal benefit for the recipient (b), adjusted for the level of genetic relatedness between the two (r). This can be expressed mathematically as $-c + b\,r > 0$, known as Hamilton's Rule (Hamilton 1964). These costs and benefits will vary depending on the nature of the trait and of the social and environmental conditions (the "ecology") experienced by the population under consideration.

Behavioural ecologists study how human behaviour adapts to ecological and social conditions. This approach does not assume that adaptive outcomes are the result of conscious cost-benefit calculations. Instead, it merely posits that cost-benefit structures embedded in the socio-ecology drive the selection or adoption of a behaviour. Accordingly, while human behavioural ecologists tend to assume that behaviour has evolved to maximize inclusive fitness, they remain largely agnostic about the proximate mechanisms responsible for the behaviour, including how the trait is inherited as well as its cognitive and psychological underpinnings (Borgerhoff Mulder 1991; Borgerhoff Mulder and Schacht 2012; Mace 2014; Nettle et al. 2013; E. A. Smith and Winterhalder 1992; Winterhalder and Smith 2000). Assuming that mechanisms do not alter optimal strategies is a methodological approach known as the "phenotypic gambit" (Grafen 1984). The freedom from worrying about mechanisms, like nature and nurture (that had stagnated ethology for some time in the 1950s and 60s), was a fruitful academic pathway. That said, a resurgence of interest in mechanism, including genetics and cultural transmission, has been encompassed in the field now for some time, both amongst behavioural ecologists studying humans and animals.

Human behavioural ecology is a framework for examining the trade-offs between various competing, fitness-relevant dimensions of life, such as the desire to maximize resources invested in one's own offspring and the need to cooperate with other kin to ensure one's livelihood. As resources tend to be limited, individuals face trade-offs even when cooperating with a range of people who are all equally related to them, which is the case between parents and children or among full siblings. For example, parents invest resources and care into offspring, and this increases the offspring's chance of survival. However, increasing investment in one offspring decreases the resources available for other offspring and for future reproduction, resulting in parent-offspring conflict and sibling competition over the optimal level of investment allocated to each child (Trivers 1974). The costs and benefits incurred by an individual as a result of a given behavioural strategy – and the patterns of parent-offspring or sibling competition that result from them – have been shown to be shaped by the local socio-ecology (see for example variation in sibling competition in different kinship systems Ji et al. 2013, 2014).

22.4Cultural evolution: clarifying the confusion between phenomenon and theory

It is sometimes suggested that behavioural ecology is unfit for the study of behaviour in our species, because it downplays or completely disregards the role of cultural evolution. As we have argued elsewhere (Micheletti, Brandl and Mace 2022), we believe that this is a misconception that stems from the unhelpful use of the term "cultural evolution" to indicate both a phenomenon – culture changing through time – and a theory to explain it – the central role played by transmission biases in shaping culture and an associated focus on cultural inheritance. Ambiguity surrounding this term is widespread both in the literature and in informal discussion, and we may even have been guilty of this ourselves.

Using the same term for both a phenomenon and a theory generates confusion between explanans and explanandum (Hempel and Oppenheim 1948). Moreover, lack of clarity regarding the term cultural evolution leads to language suggesting ecology and culture as alternative explanations for human behavioural variation (Micheletti, Brandl and Mace 2022 review three emblematic cases in recently published influential opinion pieces). Such a dichotomy is impossible: culture and adaptation to ecology are not competing forces shaping behaviour (Micheletti, Brandl and Mace 2022). Instead, in a species with advanced cultural capabilities like our own, adaptation to local ecology is an outcome of the process of cultural evolution: some cultural variants are favoured over others by cultural selection because they are adaptive in a given environment (Boyd 2018).

For these reasons, the term cultural evolution is best reserved for the phenomenon, not any one approach or associated theories (Micheletti, Brandl and Mace 2022). Approaches centred on the role of cultural transmission should not be referred to as "cultural evolution" if confusion is to be avoided. We believe that the term "cultural evolutionary theory" is also problematic because it implicitly suggests that there is only one theory that acknowledges cultural change and that its view of cultural change is the only one possible. It is beyond the scope of this chapter to suggest a better name for these approaches – and it is a task for its proponents. For simplicity, here we will use "cultural transmission approaches" to refer to them.

22.5 Human behavioural ecologists have been studying cultural evolution all along

Once this confusion is resolved, it becomes clear that human behavioural ecologists have studied cultural evolution from the very beginning. Researchers employing this approach aim primarily to explain whether and how behaviours serve an adaptive function. Most human behaviours are culturally transmitted, or at least partially influenced by transmitted culture. It follows that, in most cases, human behavioural ecologists study the cultural evolution of human behaviours. Specifically, they do so in two ways. The behaviour being studied is an aspect of culture, a culturally transmitted trait whose adoption is shaped by an individual's ecology. At the same time, culture forms part of the environment that determines the costs and benefits experienced by individuals. Human behavioural ecology makes predictions about behaviour based on biological fitness goals, regardless of whether that behaviour is transmitted genetically or culturally or a bit of both (Mace 2014; Micheletti, Brandl and Mace 2022).

A behavioural ecology perspective can inform our understanding of cultural traits that have evolved as adaptive responses to problems posed by the local environment (see Holden and Mace 2003). These cultural traits ultimately arise from the actions of individuals whose choices can be expressed in formal models of fitness trade-offs. Individuals are considered as if they were goal-directed agents trying to make optimal decisions about their inclusive fitness, who face various trade-offs while doing so. This shapes how people respond to ideas and practices generated by others, and whether they copy them or not. In other words, whether individuals adopt or reject some trait they have observed in others is driven by their inclusive fitness interests – the need to survive and reproduce, and the subsidiary needs to assist relatives, defend resources, and acquire food. Of course, not all traits have an impact on biological fitness, and some cultural traits may be widespread even though there is no clear connection with the inclusive fitness of their bearers. As a result, some have argued that there is such a thing as 'cultural fitness' that can be studied independently of biological fitness. We discuss this concept, and its relationship with biological fitness, in a Section 8. Here, we focus on cultural traits that impact inclusive fitness.

In the pursuit of their fitness goals, individuals cooperate and compete with others. As a result, the behaviour of social partners becomes part of the landscape that shapes individuals' decisions and thus the dynamics under which cultural traits evolve. For example, Lamba and

Mace (2011) found that different populations of the same small-scale society differed significantly in cooperative behaviour, and the latter was associated with demographic variables such as village population and social network size. The fact that socio-ecological factors shape behaviour does not discount the role of cultural transmission – it merely suggests that "cultural transmission produces environmentally contingent patterns of behavioural variation that are similar to those produced via the genetic transmission of behaviour" (Lamba and Mace 2011, p. 14429). Culturally evolved responses to the local environment are not static. They are flexible and shift in response to changes in that environment and the trade-offs it imposes on us (Holden and Mace 2003). Short-term shifts emerge from the behavioural flexibility of individuals and generate adaptive plasticity within generations, whereas long-term shifts transform the behaviour of many individuals across generations. This process has enabled us to adapt to the diverse environments we inhabit today.

Behavioural ecological studies try to identify the adaptive drivers behind cultural traits by employing elements of the rich behavioural ecology toolkit. This toolkit that includes: mathematical models, used to generate predictions based on inclusive fitness maximisation or to test the logical consistency of an hypothesis; experiments, manipulating characteristics of either the local environment or of the individuals themselves to tease out why certain behaviours have been selected for; naturalistic optimality studies, observational research comparing the behaviour of different individuals, which often makes use of 'natural experiments'; and cross-species comparisons using phylogenetic methods, which model how traits have co-evolved over time and identifying the ecological drivers of change from ancestral states (Davies et al. 2012; Krebs and Davies 1978).

As experimental manipulations relating to fitness are not possible in humans due to obvious ethical reasons, in order to test predictions, human behavioural ecologists rely mainly on observational studies within populations and cross-cultural comparisons (instead of cross-species comparisons, Pagel and Mace 2004). The comparative approach focuses on systematic differences between cultures that are relatively homogenous within each group, such as languages and kinship systems. When performing cross cultural comparisons, it is necessary to account for "Galton's problem", the fact that similarities between cultures may stem from shared ancestry rather than having evolved independently. Phylogenetic cultural comparative studies allow us to control for this phylogenetic association, thus solving Galton's problem (Mace and Pagel 1994). And, possibly more importantly, some phylogenetic comparative methods provide a formal model of the direction of cultural evolutionary processes that can be tested.

Observed patterns of cultural diversity show how cultures vary; however, they do not explain why they arose. Phylogenetic comparative methods help us understand how the present cultural diversity arose from ancestral states by explicitly modelling the evolutionary pathways of cultural traits (Mace and Zhang in press), i.e. Tinbergen's 'evolutionary history' (or 'phylogeny') question. They can be applied to test hypotheses regarding the co-evolution of a cultural trait and the ecological features experienced by a population, which reflects the adaptiveness of cultural traits throughout evolutionary history, i.e. Tinbergen's 'evolutionary function' question. If the cultural groups being studied evolved by descent with modifications, we can reconstruct the hierarchical descent of cultural groups using archaeological, genomic, and most often language data. Phylogenetic reconstructions of major language families in the world (e.g., Grollemund et al. 2015; Zhang et al. 2020) are available for use in cross-cultural comparative studies. Phylogenetic comparative studies have shown that the evolutionary trajectories of many cultural traits are lineage-specific and cannot be generalised to other language families (e.g., Dunn et al. 2011; Passmore and Jordan 2020). Thus, behavioural ecological thinking has a vital role to play in helping us understand cultural diversity. Fitness-based predictions have been shown empirically to be strong and have by no means been usurped by considerations of cultural transmission biases. Inclusive fitness is still the best general framework for understanding the diversity of cultural behavioural phenotypes that we see around us.

We now briefly consider three human cultural behaviours that have been successfully studied with a behavioural ecological approach: kinship, religious institutions, and witchcraft belief. These were all foundational topics of study in anthropology. We show that tools and perspectives from behavioural ecology can helps us elucidate the inclusive fitness costs and benefits of a trait to generate predictions and test them.

22.5.1. Kinship

Human kinship systems are complex social systems that regulate descent, residence, marriage, and inheritance. Most unilineal cultures described in the ethnographic record are patrilineal, meaning that descent, inheritance, or both, are derived from the father (590 societies recorded in the Ethnographic Atlas; Kirby et al. 2016; Murdock et al. 1999). However, matrilineal kinship systems, where titles or resources may derive from the mother or the maternal uncle, remain present in a substantial minority of human cultures (160 societies in the Ethnographic Atlas, the remaining 524 cultures having bilateral, ambilineal, or other ambiguous descent systems; Kirby et al. 2016; Murdock et al. 1999). The rise of human behavioural ecology stimulated a resurgence in interest in kinship, which whilst foundational in anthropology was becoming a tired topic (Mattison et al. 2019; Shenk and Mattison 2011).

One key feature of kinship – the transmission of material wealth – can be conceptualized as a form of parental investment: parents must decide which offspring to invest wealth in. Optimal decisions are considered to be those that divide resources between children in such a way as to maximize the number of grandchildren (Mace 1998). The transmission of material wealth - such as land or livestock – can support people's reproductive careers because they need resources to attract a spouse and support offspring. This line of reasoning does not equate economic with genetic inheritance. Rather, it holds that parental care (of which the inheritance of material possessions is one instance) is targeted towards particular children in ways that we can make sense of in terms of fitness trade-offs. This means that parents should preferentially transmit wealth to those offspring whose reproductive careers will benefit most from inheriting their parents' possessions. This in turn benefits the parents by providing them with more grandchildren. Matrilineal descent and transmission of wealth is favoured whenever parents benefit from investing in daughters rather than sons, and vice versa for patrilineal descent (Holden and Mace 2003). Which sex parents should invest in – and thus, which descent and inheritance system they should favour – is determined by features of the local ecology, such as the subsistence strategy they pursue, and how these interact with the constraints of biological reproduction.

Wealth usually has a greater impact on the reproductive potential of sons than it does on daughters, due to its role in enabling polygyny. In other words, in settings where polygyny is allowed, wealth enables men to attract multiple wives, which in turn enables them to have more children than they would have had in a monogamous marriage. In such environments, parents benefit from preferentially giving wealth to sons. But unlike daughters, sons also face the risk that the children born to their partners may not be their own. In mating systems where paternity uncertainty is high, parents benefit more from placing 'safe bets' in their daughters, whose offspring are guaranteed to be their biological grandchildren. If paternity certainty is held constant, then any increase in resources that have a greater positive impact on the reproductive success of males incentivizes parents to invest that resource in sons (Holden and Mace 2003). For example, among the Gabbra, a pastoralist population in Kenya, the size of a household's camel herd has a greater positive effect on the reproductive success of men than that of women due to polygyny (Mace 1996). As a result, we would expect that societies that keep livestock should invest more wealth in sons than in daughters (Hartung 1982). Confirming this prediction, phylogenetic comparative analyses of Bantu populations in sub-Saharan Africa have revealed that descent systems have co-evolved with subsistence strategies (Holden and Mace 2003). The adoption of cattle-breeding, which introduced wealth items that men could defend and monopolize for bride price payments (wealth transfers from the groom's family to the bride's), induced matrilineal societies to switch to patrilineal or mixed descent systems (Holden and Mace 2003). This result has been recently confirmed in a worldwide analysis of transitions to and away from matriliny (Shenk et al. 2019). Insights from inclusive fitness theory and parental investment can help us explain the distribution of kinship systems around the world.

22.5.2. Religious institutions

Participation in religious activities has long puzzled the evolutionary human sciences, as it is both widespread and requires considerable time and energy investments that do not always seem to contribute to fitness. No other is more costly in terms of personal fitness than lifelong celibacy, which is demanded of religious specialists in various forms in several of the major world religions, including Christianity, Buddhism, Hinduism, Jainism, and some Sufi groups within Islam (Qirko 2002). Researchers focussing on cultural transmission have argued that this clearly maladaptive practice could arise and spread because it helped enhance the credibility of these religious practitioners (Singh and Henrich 2020), and some suggest that it could further spread, together with beneficial cooperative norms, through cultural group selection (Henrich 2009; Norenzayan et al. 2016).

Alternatively, lifelong celibacy could be acceptable to parents, if inducing some children to become celibate allows them to distribute their material wealth optimally, furthering their own reproductive success. Genealogical data from Medieval and Early Modern Europe have offered some limited support for this view, showing that richer families directed more children to the seminary or to the cloister (Boone 1986) and that, in a comparison of two French noble families, the one with more celibates persisted for a longer time (Hill 1999). However, these analyses could not establish clearly whether and how this practice could be adaptive for the celibates themselves, their siblings, and their parents.

Recent work adopting a behavioural ecology approach has investigated the fitness trade-offs associated with lifelong religious celibacy, using sociodemographic data from an agropastoralist population in western China where parents sent sons to the monastery until recently (Micheletti, Ge, Zhou et al. 2022). They found that the presence of a monk is associated with more children for the monk's brother and more grandchildren for the monk's father, suggesting that the practice is indeed adaptive for them. Furthermore, men with a celibate brother have more children than men with a non-celibate one. The reproductive success of brothers of monks is similar to that of only sons; this suggests that the decision to send a son to the monastery is a way for parents to reduce sibling conflict (Micheletti, Ge, Zhou et al. 2022). In addition, parents preferentially send second or later born sons to the monastery, whereas first born sons are more likely to inherit the family household (Zhou, Ge et al. under review). This parental bias towards older offspring is in line with the results of inclusive fitness analyses (Jeon 2008) and has been observed in both industrialised and pastoralist societies (Hrdy and Judge 1993; Mace 1996).

To understand the long-term evolutionary dynamics of religious institutions, phylogenetic comparative inferences are useful where empirical records (e.g. archaeological artefacts, written history) are available. The fairly complete historical record of world religions provides a unique opportunity to reconstruct its evolutionary history on a phylogeny calibrated by timings of splitting and extinction events as recorded in the written history. A recent phylogenetic comparative study (Basava et al. 2021) tests whether intrinsic elements of religious institutions (i.e. eschatological ideologies) may influence their survival and extinction and found evidence that apocalyptic beliefs are associated with accelerated group extinction even after phylogenetic associations are controlled; it also found substantial evidence that apocalyptic beliefs co-evolved with revolutionary violence, while reincarnation beliefs were evolutionarily stable in peaceful groups. In both cases, violence precedes the emergence of the belief, which suggests that conditions which generate revolutionary violence changed beliefs rather than beliefs generated violence.

22.5.3. Witchcraft belief

Belief in witchcraft, or the idea that some individuals harm others through supernatural means, is another cultural trait often viewed as an irrational superstition. Actions intended to combat witchcraft are often undertaken to cure illness, despite being scientifically ineffective. Such beliefs may be maintained in populations through cultural evolutionary processes based on frequency dependent learning among other things (Tanaka et al. 2009). Yet witchcraft beliefs may have an adaptive function: accusations might be a mechanism for nullifying competitors, which benefits accusers by allowing greater access to resources or separation from costly relationships (Douglas 1991; Mace et al. 2018). Of course, behavioural ecologists do not necessarily assume that witchcraft accusations will benefit the survival or reproductive careers of accusers directly. But by eliminating competitors, accusers may gain resources (such as material wealth) that can support their survival and reproduction in the long term. The risk of retaliation or reputational damage to accusers (Jordan et al. 2016) is mitigated by giving targets the negative 'tag' of 'witch'. This can justify mistreatment (Singh 2021) and protect accusers' own reputations. Empirical studies do not support the hypothesis that witchcraft accusations primarily target uncooperative individuals (Mace et al. 2018). A link with interpersonal competition is suggested by the correlation between increases in witchcraft murders and trials with intensifying environmental hardship and resource scarcity both in modern-day Tanzania and in early modern Europe (Miguel 2005; Oster 2004).

Studies also suggest that witchcraft belief co-evolves with competition for material wealth, with stronger beliefs in societies depending on agricultural subsistence than among relatively egalitarian hunter-gatherers (Guenther 1992; Koning 2013). Accusations seem to be more frequent in the context of relationships that produce conflict and competition between individuals. Peacey et al. (under review) built a dataset of ethnographic texts from societies in sub-Saharan Africa containing evidence of witchcraft accusations to explore this possibility. They found that different forms of competitive relationship may determine who is likely to be accused. Men were more often accused by their relatives and by unrelated individuals, resulting from male-male patterns of competition for resources such as bride price. Women were more likely to be accused in the context of relationships with their affinal kin, resulting from factors such as competition with co-wives in polygynous marriages for reproductive resources from their husbands. There were also cases of husbands accusing their wives (and a very small number where the situation was reversed), in scenarios that appear to stem from underlying inter-sexual conflict. But accusations of men, by men, were the most common form of accusation in this study, perhaps indicating male-male competition was the predominant form within the socio-ecology of these cultures (Peacey et al. under review).

In other societies, the direction of competition may vary, suggesting why the sex and categories of individuals most likely to be accused of witchcraft differs across cultures. For example, early modern Europe had far more female witches and far fewer witchcraft accusations between kin than African cultures (Hutton 2004). But despite the variation in who is targeted, witchcraft accusations have enough recurrent features to suggest they are a flexible adaptation that occurs in diverse locations. At their basic level they mostly occur between individuals who are competing or in conflict, which should be viewed as separate from the role of large-scale institutional involvement which occurred within the European witch trials (Sharpe 1996). This shows that socio-ecological factors commonly explored by behavioural ecologists – such as resource availability and patterns of cooperation and coopetition within societies – shape the frequency and direction of witchcraft accusations.

22.5.4. Maladaptive behaviours and the role of transmission dynamics

The examples above have shown that generating functional predictions based on inclusive fitness has helped us explain the variation and distribution of a range of cultural phenotypes around the world. However, this does not imply that behaviour is always optimal. Maladaptive behaviour can arise in populations for multiple reasons. When the environment changes, so do cost-benefit scenarios for individuals, and adjusting to a new optimum can take time. This is a well-documented phenomenon in evolutionary biology, known as adaptive lag, which occurs in genetic evolution. Cultural behaviours can show something comparable to adaptive lag. For example, Tibetan herders express a preference for sons, but measures of infant feeding, interbirth interval and results of a gift-giving game suggest that they invest more heavily in daughters (Du and Mace 2018). In this society, it is possible that cultural preferences are 'out of date', as the traditional role of males (who were formerly heavily focussed on warfare and herd defence) is being forced to change fast. How exactly mothers make strategic decisions about parental investment is not known. It could be 'learning biases' or their own strategic assessments of costs and benefits (or a combination of both). It should be noted that, despite the prominence accorded to it in evolutionary psychology (see Tooby and Cosmides 1990), adaptive lag is not atypically large in humans when compared to other animals (Laland and Brown 2006). Other reasons for maladaptive outcomes include power asymmetries; these may force individuals to follow strategies that are not in their own best interest due to coercion or the costs of punishment from stronger parties in their society. Evolutionary stable strategies may depend on starting conditions, leaving some behaviours marooned in adaptive peaks or valleys, unable to cross to a higher fitness peak. Lastly, in small populations, cultural drift may prevent the attainment of optimal solutions.

Cultural transmission dynamics can sometimes prevent the realisation of inclusive fitness interests, leading to maladaptive behaviours; more empirical research is needed to establish when this is indeed the case (the demographic transition from high to low fertility is one candidate). In recent years, some have argued that, for this reason, the behavioural ecological approach in its current form cannot study adaptation satisfactorily and that considering cultural transmission dynamics is essential (Borgerhoff Mulder 2013; Creanza et al. 2017). We disagree. Fitness-based models and theories can make meaningful predictions about cultural diversity precisely because they do not consider dynamics. In most cases, little is known about how precisely a specific trait has been transmitted over time and perhaps very little can ever be known, as multiple mechanisms are likely to be acting at the same time. The alternative is to do comparative statics, that is generate predictions about the expected evolutionary stable strategies under different conditions, and then test these predictions against data (cf. Frank 1998; Hammerstein 1996). This has proven invaluable in genetic social evolution theory, where often very little is or can be known about the genetic architecture of a trait. The study of the evolution of cultural traits can benefit from this approach too.

To characterise any cultural (or other) behaviour as maladaptive, one does need a model of what the adaptive behaviour would look like. Otherwise, the term ‘maladaptive’ is not clear. The study of design through fitness, freed from a study of transmission, remains a fruitful approach for studying behaviour, including cultural traits.

22.6 Organising research: Tinbergen’s four questions about behaviour

Cultural change can be studied from a range of different perspectives. These include psychology and cognitive science, population-genetics-inspired models and epidemiological-style models. We suggest that Tinbergen’s (1963) four questions about behavioural evolution still offer a useful framework for organising research in the evolutionary human sciences. The four questions concern:

- **Mechanism:** the physical apparatus that causes an individual to manifest a behaviour, including sensory, nervous, hormonal, and skeletal-muscular systems. This also includes the cognitive and psychological traits that enable behaviour.
- **Ontogeny:** how the behaviour changes throughout the life course of the individual, with special attention to how it is formed during development, and this may include how behaviour is shaped through learning, and 'nature vs nurture', genetic vs environmental influences.
- **Function:** what a behaviour is 'good for', whether it is biologically adaptive and if so in what way.
- **Phylogeny:** the evolutionary history of the behaviour, how it evolved in the ancestors of individuals currently expressing it.

These questions were originally advanced by Tinbergen (1963) to investigate the evolution of behaviour in non-human animals, under the belief that behavioural traits were genetically controlled. While this is true for most animal behaviours (Davies et al. 2012), since then it has been found that cultural transmission plays a role in many animal species (Schuppli and Schaik 2019). However, this has not decreased the utility of Tinbergen's approach to organising research and knowledge about animal behaviour. Similarly, these questions remain valid for human behaviour that are – at least partly – culturally transmitted.

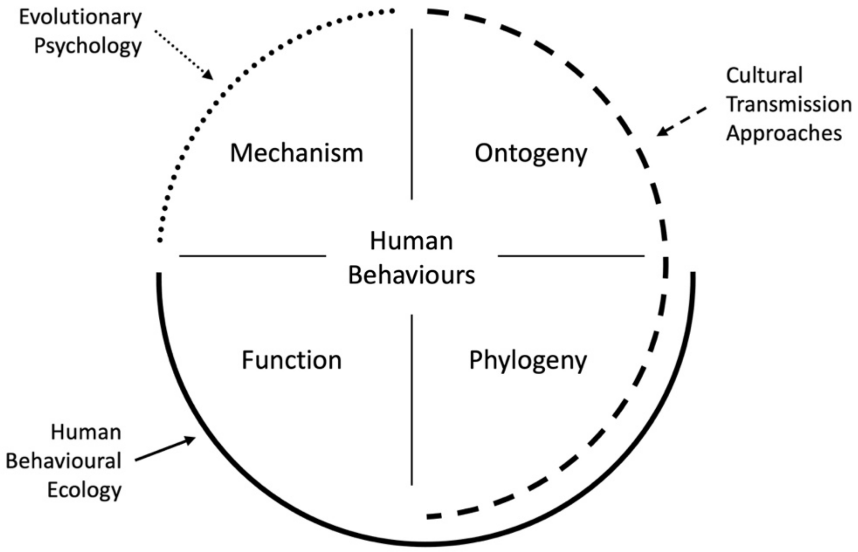


Figure 1*: Different evolutionary approaches ask different questions about human behaviours in Tinbergen's (1963) schema: human behavioural ecology (solid line) is concerned with function and phylogeny; the cultural transmission approaches (dashed line) are concerned with ontogeny and phylogeny; evolutionary psychology (dotted line) is concerned with mechanism. Diagram after Micheletti, Brandl and Mace (2022).*

Tinbergen's four questions about behaviour are not mutually exclusive; in fact, they are complementary. Since they consider different levels of explanation, answering all four is necessary to obtain a full understanding of a given trait. Consider for example the use of wild plants, especially ones with medicinal properties, by human populations. When we say that humans avoid certain plants because of innate behavioural avoidance strategies and/or social learning rules exhibited since early infancy, we are answering a question about cognitive mechanisms (Elsner and Wertz 2019; Wertz 2019; Wertz and Wynn 2014; Włodarczyk et al. 2018). When we say that hunter-gatherers use certain medicinal plants because they have learnt foraging skills through a combination of vertical, oblique, and horizontal transmission (Hewlett et al. 2011; Hewlett and Cavalli-Sforza 1986; Lozada et al. 2006), we are answering a question about ontogeny, i.e. how the behaviour develops during the individual's lifetime. When we say that hunter-gatherers use medicinal plants because it improves the physical health of their children (Salali et al. 2016), we are addressing a question about function. Finally, when we try to reconstruct cultural phylogenies of plant use by different human groups to identify ancestral states, correlated evolution and the drivers of similarities in ethnofloras, we are asking a question about the phylogenetic history of this trait (Teixidor-Toneu et al. 2018). The three evolutionary approaches to human behaviour can all contribute to increasing our understanding of human behaviour, but it is important to realise that different approaches tackle different questions in Tinbergen's (1963) schema (see Figure 1).

22.7Difficulties with cultural transmission approaches to cultural evolution

Approaches to cultural evolution that focus solely on social learning mechanisms and transmission biases have sometimes been equated with cultural evolution itself. As Micheletti, Brandl and Mace (2022) have argued (see also Section 4), this conflates the empirical phenomenon of cultural change with a particular tradition of theorizing and a set of specific methods to investigate this phenomenon. In this section, we briefly explore some difficulties with these approaches. We first discuss some semantic confusions that spring from the model of the human mind adopted by cultural transmission researchers; and we then turn to social learning biases, which often form the conceptual centrepiece of these approaches.

Proponents of these approaches tend to work with a model of the human mind that conceives of humans first and foremost as social learners (Powers et al. 2021). This may explain why they sometimes treat mechanisms as if they were functional explanations – for example, by pointing to mechanistic constructs such as norm psychology as an explanation for cooperation that either complements or competes with kin selection (Henrich and Muthukrishna 2021). Proponents of this view have argued that 'genetic evolutionary mechanisms' like kin-based altruism cannot explain the high levels of cooperation observed in human societies, and that only gene-culture co-evolution can (Henrich and Muthukrishna 2021). In their account, cultural evolution 'favoured' the emergence of prosocial norms in ancestral environments, people began to sanction norm violators, and this created genetic selection pressures for psychological dispositions that repressed aggression and enabled norm adherence (Henrich and Muthukrishna 2021). We have pointed out elsewhere that cultural evolution cannot ‘favour’ a particular outcome – only some form of selection can (Micheletti, Brandl and Mace 2022). Sanctioning aggressors who violate prosocial norms is perfectly compatible with kin-based altruism, as doing so would have protected the interests of the punisher’s spouses, children, and other relatives, especially in small foraging bands. Clearly, this account does not provide an alternative to kin selection but instead describes a possible instance thereof.

The underlying confusion may come from the fact that kin selection is sometimes described as an evolutionary mechanism. This is an unfortunate rhetorical convention that most researchers in the evolutionary social sciences are guilty of, including behavioural ecologists. It is unhelpful because it blurs the line between functional pathways, which explain how a trait generates fitness benefits, and mechanisms in Tinbergen’s (1963) sense of that term, which address how a behaviour ‘works’ in real time. Kin selection is a functional pathway. In contrast, we believe that norm psychology is a mechanism in Tinbergen’s (1963) sense (in this case, a set of psychosocial dispositions that motivate us to act in a certain way) but not a functional pathway. As such, norm psychology cannot ‘complement’ kin selection as an explanation for the evolutionary function of cooperation, that is, norm psychology is not the answer to the question: ‘why was cooperation selected for?’. Instead, norm psychology can ‘complement’ kin selection by illuminating how cooperation is represented in the minds of individuals – the way they think and feel about cooperative activities they are participating in. In other words, ‘norm psychology’ offers an explanation at the level of the mechanism – not of the function.

The mechanism-first approach to empirical investigations focusses heavily on *how* cultural traits are replicated (i.e. how they are transmitted). It is less concerned with *why* cultural traits and institutions arise, persist, and dissolve. These are equally important evolutionary questions that can be answered empirically (see review in Zhang and Mace 2021). Frequency-dependent fitness costs and benefits are key to understanding both the origin and the loss of cultural diversity. Evolutionary models informed by empirical data (Ji et al. 2016) have shown that cultural norms of marital residence can evolve as a frequency-dependent strategy in response to changing fitness payoffs and starting conditions. These incentives are necessary for our understanding of cultural change, which requires that people diverge from pre-existing norms rather than upholding them.

Some theorists talk about the possibility of direct or pay-off bias, or copying the ‘best’ strategy (Boyd and Richerson 1985). However, that begs the question of how we define ‘best’. If we do so in terms of fitness, then we are back to behavioural ecology. Behavioural ecologists often assume that these pay-offs can be a proxy for fitness (Davies et al. 2012; Krebs and Davies 1978). Behavioural economics studies using lab-in-the-field experiments also often assume that such a correspondence exists (e.g., Wu et al. 2015). However, if the pay-off is more short-term and immediate, then maximizing these pay-offs may not necessarily correspond to maximizing fitness. Others have already pointed out that models of human behaviour that assume us to be ‘rational strategizers’ (who always strive to maximize their own material pay-offs, for example immediate financial rewards) are not identical with those used by behavioural ecologists (who assume that agents maximize inclusive fitness) (Powers et al. 2021).

Payoff-insensitive biases might explain cultural behavioural change in environments where the ecology or other factors have changed fast, and where behavioural ecology approaches may not apply (we address this point in Sections 5.4 and 8). Some theorists have proposed that some biases – most notably conformity and prestige bias – generate adaptive outcomes even when followed blindly, and that they are therefore key to cultural adaptation. In other words, learning biases enable us to ‘sleepwalk’ into adaptive outcomes, like a car driving with the headlights off. However, humans do not blindly copy others but instead forage for information and apply common-sense heuristics about who and what may be useful (Bellamy et al. 2021; Deffner et al. 2020; D. Smith 2020). While humans have occasionally been portrayed as enthusiastic over-imitators (to the point of blind conformity), recent work shows that the tendency to over-imitate (copying causally irrelevant actions) is very flexible, short-term, and contextual (Keupp et al. 2015; Kline et al. 2020; Rawlings et al. 2019). This suggests that the tendency to conform is not a coherent ‘bias’. Similar criticisms have already been made of prestige bias. Circular definitions of prestige mean that many proposed instances of prestige bias are in fact also consistent with generic goal-directed action (Chellappoo 2020). It appears that human cognitive systems have not been selected to execute a narrow set of biases or stable learning rules, but instead have been selected for a domain-general Bayesian ability to adjust their learning as they go along (Stout 2013; Whiten 2015). Humans modify their tools, techniques, and knowledge base in response to experience. It is unclear how transmission biases would generate the high level of plasticity we observe in human social learning.

Transmission mechanisms, such as teaching and imitation, and biases, such as prestige and conformity bias, are certainly compelling issues worthy of exploration. However, if the phenomenon of cultural evolution is reduced to this process, that limits our understanding of cultural diversity. With social learning processes showing few general rules, predicting cultural diversity from the ground up, namely from the mechanisms of social transmission, is going to be a very hard task (Micheletti, Brandl and Mace 2022). Social psychologists have been trying to predict behaviour from mechanistic constructs such as priming (where subtle cues in a person's environment are said to change their behaviour) for a long time. But many key findings in social psychology have failed to replicate – for example, the idea that priming people with religious references increases prosocial behaviour and discourages free riding. While an earlier meta-analysis concluded that religious priming effects are robust (Shariff et al. 2015), re-analyses have cast doubt on this conclusion (van Elk et al. 2015). Furthermore, many classic studies on religious priming fail to hold up in large, pre-registered replications (see review in Ross 2021). These challenges may be rooted in the weakness of the original predictive hypotheses; this is compounded by a 'theory crisis' where many prominent psychological constructs lack validity (Eronen and Bringmann 2021). A firm theoretical framework for hypothesis-building may be a useful bulwark against the replication crisis; and the predictions offered by taking a behavioural ecological approach may help in that regard.

22.8 Moving forward

Approaches focussed on fitness and approaches focused on cultural transmission can fruitfully coexist and complement each other. As we have already mentioned, we believe that the framework offered by Tinbergen's (1963) four questions still offer a useful guide. The three schools of thought that characterised evolutionary studies of human behaviour are a useful description of researchers' disciplinary backgrounds and differences in perspectives on human behaviour when the field started developing; but they do not necessarily reflect the current state of the field, as it is moving towards greater integration. However, a true synthesis has not yet been reached. A recently proposed distinction between two cultural selection processes may contribute to achieving greater clarity.

Birch (2017) has argued that two distinct selective processes operate in cultural evolution (see also Birch and Heyes 2021; Micheletti 2020; D. Smith 2020). The first is Cultural Selection

1 (CS1), that is selection of cultural variants based on the biological fitness (number of biological descendants) of their bearers; in other words, cultural traits are selected for because they increase the biological fitness of the people who have these traits (Birch 2017). The other process is Cultural Selection 2 (CS2), that is selection of cultural variants based on the 'cultural fitness' of their bearers (measured in the number of learners) (Birch 2017). In other words, Cultural Selection 2 operates through people's ability to generate 'cultural offspring', that is their capacity to attract more learners or 'apprentices' to transmit a trait to. Similar to biological fitness, 'cultural fitness' is a property of the individual – but it is defined with respect to a specific cultural trait. Higher 'cultural fitness' accrues to individuals who transmit a given cultural variant to more people (Birch 2017; Birch and Heyes 2021).

CS1 is not equivalent to genetic natural selection, because the former acts on cultural variants, whereas the latter on cultural variants (Birch 2017; Birch and Heyes 2021). However, the two share the same currency, biological fitness. Therefore, in species that have cultural transmission – like our own – genetic natural selection is not the only force that can generate adaptation: CS1 can also lead to it. As both CS1 and CS2 concern cultural variants, the distinction between the two is not equivalent to that between genetic evolution and cultural evolution. It is not a reformulation of gene-culture coevolution, but neither is it an alternative, as this framework has undoubtedly explanatory power in some cases (Birch and Heyes 2021).

Some theorists have proposed that CS1 is slower and emerged before CS2, and that the latter has become the dominant force in human cultures over time (Birch 2017; Birch and Heyes 2021). While this account has potential strengths, we wonder whether both processes might still be operating, albeit on different traits, or perhaps to different degrees on the same trait, in some cases. Future work is needed to clarify the relationship between CS1 and CS2 – and theoreticians should develop new ways to model the two processes. But if the distinction between the two is proven to be solid, it could help us organise our thinking and research about cultural evolution.

CS1 should constrain cultural evolution most dramatically for traits that are directly relevant to survival and reproduction. In contrast, cultural traits that do not obviously influence inclusive fitness (such as decorative patterns on pottery) may evolve in ways that are more decoupled from fitness incentives imposed by the local ecology, that is they may evolve under the action of CS2. Accordingly, group-level cultural patterns are not reducible to transmission biases, specific transmission mechanisms or network structures, at least when CS1 is at work – just like the distribution of physiological traits among organisms is not reducible to the biochemical mechanisms involved in copying and recombining DNA. Another way to put this may be that when human behavioural ecologists examine cultural evolution, they do so with a focus on CS1. On the other hand, cultural evolution may indeed be reducible to these transmission biases when CS2 is at work, i.e. the focus is on traits that are neutral with respect to inclusive fitness. This may be the case for kinship terminologies, for example.

We have already discussed how decisions about the inheritance of material wealth have direct repercussions for inclusive fitness; as a result, we can use principles derived from human behavioural ecology to make predictions about the cultural evolution of inheritance systems (see Section 5.1). Other features of human kinship may be less relevant for reproduction. For example, human societies have about six distinct terminological systems, including for example Hawai'ian (where all the parents' brothers and sisters are addressed as mother and father) and Iroquois (where the parents' same-sex siblings are addressed as mother and father but opposite-sex siblings are addressed by a different term) (see Morgan 1871).

Anthropologists have hypothesized that the characteristics of these terminologies (such as the differentiation between cross-cousins and parallel cousins, i.e. the children of mother's sister/father's brother and mother's brother/father's sister) have co-evolved with features of the social structure such as marriage taboos (for an overview and a critical perspective see Coult 1965). Comparative analyses using language phylogenies have found little support for most of these hypotheses (Passmore and Jordan 2020). Some have already pointed out that 'kinship terms are not kinship' (Bloch 2010); in other words, terminological systems are only a small part of the broader phenomenon of kinship (Bloch 2010). While linguistic conventions may have frequency-dependent consequences for biological fitness (people need to be understood to communicate with their peers), which kin terminology a community of speakers settles on should not influence the reproductive success of individuals within that community, at least not directly. As a result, and unlike inheritance systems, the cultural evolution of terminologies can operate independently of the fitness incentives imposed by the local environment. On the other hand, the structure of existing terminologies may have cognitive and communicative advantages over unrealized alternatives (for example, by being informative but also easy to use and remember) (see Jones 2010 on the 'grammar of kinship'). The known terminological systems may thus have emerged under CS2 rather than CS1. Again, we believe that this aspect of human kinship is where approaches that focus on cognitive mechanisms can contribute the most to our understanding of cultural evolution.

While the CS1/CS2 distinction may be a useful tool to aid thinking, it is not always clear if such a distinction can be found empirically, and some conceptual challenges remain. In particular, if CS2 selects for 'cultural fitness' (as opposed to biological fitness), theorists should address whether CS1 and CS2 differ in what counts as a mechanism, and where Cultural Selection 2 draws the line between mechanism and function. Theorists should also try to clarify the idea of 'cultural fitness' (Birch 2017; El Mouden et al. 2014; Ramsey and De Block 2017). For example, the primary measure of a person's 'cultural fitness' is the popularity of their cultural products; this may lend itself to circular reasoning when we try to understand why some individuals are copied more frequently than others. Others have already pointed out that experiments purporting to show evidence of prestige bias often rely on circular definitions of prestige (by treating the act of copying a particular person as being itself evidence of prestige Chellappoo 2020). There is potential for developing more formal models of CS1, with particular attention to how shared cultural ancestry can lead to (cultural) relatedness (Allison 1992; Birch 2017; Micheletti 2020). This would help us integrate research on social learning processes with approaches focused on biological fitness (Micheletti 2020).

Future work could also explore potential trade-offs between biological fitness and 'cultural fitness', or how successful we are at spreading our cultural products through society (El Mouden et al. 2014; Richerson and Boyd 2005). This could be helpful for our understanding of apparently maladaptive cultural phenomena resulting in low fertility. A prominent and far-reaching example of this is the demographic transition that is occurring in nearly all regions of the world, which can be explained partially but not entirely from various evolutionary perspectives (Borgerhoff Mulder 1998; Colleran 2016; Mace 1998). Here is a clear case where selection for high fertility seems to be in direct opposition to cultural selection for high educational attainment (Kong et al. 2017). However, the long and mired history of the study of a distinction between 'nature' and 'nurture', which has not been a productive academic path, warns us that a distinction between cultural and biological selection is also going to be difficult to make.

22.9 Conclusion

The term ‘cultural evolution’ is often used for both a phenomenon and a specific set of theories and approaches used to study that phenomenon, and this has created confusion in the literature. We believe that, to avoid further confusion, the term cultural evolution is best reserved to identify a phenomenon: culture changing over time. Human behavioural ecology and perspectives focussed on mechanisms such as cognitive traits and transmission biases are both approaches with which to study cultural evolution. The toolkits of different disciplines such as experimental psychology (e.g., social learning experiments) and those of behavioural ecology address different questions about the evolution of cultural traits, working on different levels of explanation. Tinbergen’s (1963) framework remains relevant in helping us untangle these levels of explanation. For example, we may say that inheritance systems and traditions of sending children to monastic institutions are maintained through vertical transmission: people learn these cultural traits from their parents and then apply them to their own children. This statement captures ontogeny rather than function. Questions about function are captured in the kinds of explanations advanced by behavioural ecologists, for example, that preferentially transmitting resources to sons or sending some children to monasteries benefits the fitness interests of their parents. While not all cultural traits are relevant to biological fitness, we argue that models informed by inclusive fitness considerations help us understand cultural diversity, and these considerations are ignored at some risk.

Acknowledgements

We thank Mark Dyble, members of the Human Evolutionary Ecology Group at UCL Anthropology, Karthik Panchanathan, Daniel Nettle, Philippe Huneman and Thomas Reydon for useful comments and discussion.

References

Allison, P. D. (1992). Cultural Relatedness Under Oblique and Horizontal Transmission Rules. *Ethology and Sociobiology*, *13*(3), 153–169.

Basava, K., Zhang, H., & Mace, R. (2021). A phylogenetic analysis of revolution and afterlife beliefs. *Nature Human Behaviour*, *5*(5), 604–611. https://doi.org/10.1038/s41562-020-01013-4

Bellamy, A., McKay, R., Vogt, S., & Efferson, C. M. (2021). Social-learning strategies are flexible and depend upon the value of social information (talk). Presented at the Annual Meeting, European Human Behaviour and Evolution Association, Krakow, 2021.

Birch, J. (2017). *The Philosophy of Social Evolution*. Oxford, UK: Oxford University Press.

Birch, J., & Heyes, C. (2021). The cultural evolution of cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1828), 20200051. https://doi.org/10.1098/rstb.2020.0051

Bloch, M. (2010). Kinship terms are not kinship. *Behavioral and Brain Sciences*, *33*(5), 384–384. https://doi.org/10.1017/S0140525X10001949

Boone, J. L. (1986). Parental Investment and Elite Family Structure in Preindustrial States: A Case Study of Late Medieval-Early Modern Portuguese Genealogies. *American Anthropologist*, *88*(4), 859–878. https://doi.org/10.1525/aa.1986.88.4.02a00050

Borgerhoff Mulder, M. (1991). Human behavioural ecology: studies in foraging and reproduction. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural Ecology: An Evolutionary Approach* (2nd edition.). Oxford,UK: Blackwell.

Borgerhoff Mulder, M. (1998). The demographic transition: are we any closer to an evolutionary explanation? *Trends in Ecology & Evolution*, *13*(7), 266–270. https://doi.org/10.1016/S0169-5347(98)01357-3

Borgerhoff Mulder, M. (2013). Human behavioral ecology—necessary but not sufficient for the evolutionary analysis of human behavior. *Behavioral Ecology*, *24*(5), 1042–1043. https://doi.org/10.1093/beheco/ars223

Borgerhoff Mulder, M., & Schacht, R. (2012). Human Behavioural Ecology. In *eLS*. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0003671.pub2

Boyd, R. (2018). *A different kind of animal: how culture transformed our species*. Princeton, NJ: Princeton University Press.

Boyd, R., Borgerhoff Mulder, M., & Richerson, P. J. (1997). Are cultural phylogenies possible? In P. Weingart, S. D. Mitchell, P. J. Richerson, & S. Maasen (Eds.), *Humans by Nature*. Milton Park, UK: Routledge.

Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago, IL: The University of Chicago Press.

Carneiro, R. L. (2003). *Evolutionism in Cultural Anthropology: A Critical History*. Boulder, CO: Westview Press.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural Transmission and Evolution*. Princeton, NJ: Princeton University Press.

Chellappoo, A. (2020). Rethinking prestige bias. *Synthese*, *198*(9). https://doi.org/10.1007/s11229-020-02565-8

Colleran, H. (2016). The cultural evolution of fertility decline. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1692), 20150152. https://doi.org/10.1098/rstb.2015.0152

Coult, A. D. (1965). Terminological correlates of cross-cousin marriage. *Bijdragen tot de taal-, land- en volkenkunde / Journal of the Humanities and Social Sciences of Southeast Asia*, *121*(1), 120–139. https://doi.org/10.1163/22134379-90002972

Creanza, N., Kolodny, O., & Feldman, M. W. (2017). Cultural evolutionary theory: How culture evolves and why it matters. *Proceedings of the National Academy of Sciences*, *114*(30), 7782–7789. https://doi.org/10.1073/pnas.1620732114

Currie, T. E., Greenhill, S. J., Gray, R. D., Hasegawa, T., & Mace, R. (2010). Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature*, *467*(7317), 801–804. https://doi.org/10.1038/nature09461

Darwin, C. (1859). *On the Origin of Species*. London: John Murray.

Darwin, C. (1877). *The Descent of Man* (2nd edition.). London: John Murray.

Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An Introduction to Behavioural Ecology* (4th Edition.). Hoboken, NJ: Wiley.

Dawkins, R. (1976). *The Selfish Gene*. Oxford, UK: Oxford University Press.

Deffner, D., Kleinow, V., & McElreath, R. (2020). Dynamic social learning in temporally and spatially variable environments. *Royal Society Open Science*, *7*(12), 200734. https://doi.org/10.1098/rsos.200734

Douglas, M. (1991). Witchcraft and Leprosy: Two Strategies of Exclusion. *Man*, *26*(4), 723. https://doi.org/10.2307/2803778

Driscoll, C. (2017). The Evolutionary Culture Concepts. *Philosophy of Science*, *84*(1), 35–55. https://doi.org/10.1086/688934

Du, J., & Mace, R. (2018). Parental investment in Tibetan populations does not reflect stated cultural norms. *Behavioral Ecology*, *29*(1), 106–116. https://doi.org/10.1093/beheco/arx134

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, *473*(7345), 79–82. https://doi.org/10.1038/nature09923

El Mouden, C., André, J.-B., Morin, O., & Nettle, D. (2014). Cultural transmission and the evolution of human behaviour: a general approach based on the Price equation. *Journal of Evolutionary Biology*, *27*(2), 231–241. https://doi.org/10.1111/jeb.12296

Elsner, C., & Wertz, A. E. (2019). The seeds of social learning: Infants exhibit more social looking for plants than other object types. *Cognition*, *183*, 244–255. https://doi.org/10.1016/j.cognition.2018.09.016

Eronen, M. I., & Bringmann, L. F. (2021). The Theory Crisis in Psychology: How to Move Forward. *Perspectives on Psychological Science*, *16*(4), 779–788. https://doi.org/10.1177/1745691620970586

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford, UK: The Clarendon Press.

Frank, S. A. (1998). *Foundations of Social Evolution*. Princeton, NJ: Princeton University Press.

Grafen, A. (1984). Natural selection, kin selection and group selection. In *Behavioural ecology: an evolutionary approach* (2nd edition.). Oxford, UK: Blackwell Publishing.

Grafen, A. (2007). The formal Darwinism project: a mid-term report. *Journal of Evolutionary Biology*, *20*(4), 1243–1254. https://doi.org/10.1111/j.1420-9101.2007.01321.x

Grafen, A. (2014). The formal darwinism project in outline. *Biology & Philosophy*, *29*(2), 155–174. https://doi.org/10.1007/s10539-013-9414-y

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C., & Pagel, M. (2015). Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences*, *112*(43), 13296–13301. https://doi.org/10.1073/pnas.1503793112

Guenther, M. G. (1992). “Not a Bushman Thing.” Witchcraft among the Bushmen and Hunter-Gatherers. *Anthropos*, *87*(1./3.), 83–107.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, *7*(7), 1–52.

Hammerstein, P. (1996). Darwinian adaptation, population genetics and the streetcar theory of evolution. *Journal of Mathematical Biology*, *34*(5–6), 511–532. https://doi.org/10.1007/BF02409748

Hartung, J. (1982). Polygyny and Inheritance of Wealth. *Current Anthropology*, *23*(1), 1–12. https://doi.org/10.1086/202775

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, *15*(2), 135–175.

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, *53*(1), 3–35. https://doi.org/10.1016/S0167-2681(03)00094-5

Henrich, J. (2009). The evolution of costly displays, cooperation and religion. *Evolution and Human Behavior*, *30*(4), 244–260. https://doi.org/10.1016/j.evolhumbehav.2009.03.005

Henrich, J., & Muthukrishna, M. (2021). The Origins and Psychology of Human Cooperation. *Annual Review of Psychology*, *72*(1), 207–240. https://doi.org/10.1146/annurev-psych-081920-042106

Hewlett, B. S., & Cavalli-Sforza, L. L. (1986). Cultural Transmission among Aka Pygmies. *American Anthropologist*, *88*(4), 922–934. https://doi.org/10.1525/aa.1986.88.4.02a00100

Hewlett, B. S., Fouts, H. N., Boyette, A. H., & Hewlett, B. L. (2011). Social learning among Congo Basin hunter–gatherers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1168–1178. https://doi.org/10.1098/rstb.2010.0373

Hill, E. (1999). Lineage interests and nonreproductive strategies: An evolutionary approach to medieval religious women. *Human Nature*, *10*(2), 109–134. https://doi.org/10.1007/s12110-999-1011-z

Holden, C. J., & Mace, R. (2003). Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1532), 2425–2433. https://doi.org/10.1098/rspb.2003.2535

Hoppitt, W., & Laland, K. N. (2013). *Social Learning. An Introduction to Mechanisms, Methods and Models*. Princeton, NJ: Princeton University Press.

Hrdy, S. B., & Judge, D. S. (1993). Darwin and the puzzle of primogeniture. *Human Nature*, *4*(1), 1–45. https://doi.org/10.1007/BF02734088

Hutton, R. (2004). Anthropological and Historical Approaches to Witchcraft: Potential for a New Collaboration? *The Historical Journal*, *47*(2), 413–434.

Jeon, J. (2008). Evolution of parental favoritism among different-aged offspring. *Behavioral Ecology*, *19*(2), 344–352. https://doi.org/10.1093/beheco/arm136

Ji, T., Wu, J.-J., He, Q.-Q., Xu, J.-J., Mace, R., & Tao, Y. (2013). Reproductive competition between females in the matrilineal Mosuo of southwestern China. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1631), 20130081. https://doi.org/10.1098/rstb.2013.0081

Ji, T., Xu, J.-J., & Mace, R. (2014). Intergenerational and Sibling Conflict Under Patrilocality. *Human Nature*, *25*(1), 66–79. https://doi.org/10.1007/s12110-013-9188-6

Ji, T., Zheng, X.-D., He, Q.-Q., Wu, J.-J., Mace, R., & Tao, Y. (2016). Kinship as a frequency dependent strategy. *Royal Society Open Science*, *3*(2), 150632. https://doi.org/10.1098/rsos.150632

Jones, D. (2010). Human kinship, from conceptual structure to grammar. *Behavioral and Brain Sciences*, *33*(5), 367–381. https://doi.org/10.1017/S0140525X10000890

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, *530*(7591), 473–476. https://doi.org/10.1038/nature16981

Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social Learning Strategies: Bridge-Building between Fields. *Trends in Cognitive Sciences*, *22*(7), 651–665. https://doi.org/10.1016/j.tics.2018.04.003

Keupp, S., Behne, T., Zachow, J., Kasbohm, A., & Rakoczy, H. (2015). Over-imitation is not automatic: Context sensitivity in children's overimitation and action interpretation of causally irrelevant actions. *Journal of Experimental Child Psychology*, *130*, 163–175. https://doi.org/10.1016/j.jecp.2014.10.005

Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H.-J., et al. (2016). D-PLACE: A Global Database of Cultural, Linguistic and Environmental Diversity. *PLOS ONE*, *11*(7), e0158391. https://doi.org/10.1371/journal.pone.0158391

Kline, M. A., Gervais, M. M., Moya, C., & Boyd, R. (2020). Irrelevant-action imitation is short-term and contextual: Evidence from two under-studied populations. *Developmental Science*, *23*(3). https://doi.org/10.1111/desc.12903

Kong, A., Frigge, M. L., Thorleifsson, G., Stefansson, H., Young, A. I., Zink, F., et al. (2017). Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences*, *114*(5), E727–E732. https://doi.org/10.1073/pnas.1612113114

Koning, N. (2013). Witchcraft Beliefs and Witch Hunts: An Interdisciplinary Explanation. *Human Nature*, *24*(2), 158–181. https://doi.org/10.1007/s12110-013-9164-1

Krebs, J. R., & Davies, N. B. (1978). *Behavioural ecology : an evolutionary approach*. Oxford, UK: Blackwell Publishing.

Kroeber, A., & Parsons, T. (1958). The Concepts of Culture and of Social System. *American Sociological Review*, *23*(1), 582–583.

Kuper, A. (1999). *The Anthropologist's Account.* Cambridge, MA: Harvard University Press.

Laland, K. N. (2004). Social learning strategies. *Learning & Behavior*, *32*(1), 4–14.

Laland, K. N., & Brown, G. R. (2006). Niche construction, human behavior, and the adaptive-lag hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *15*(3), 95–104. https://doi.org/10.1002/evan.20093

Laland, K. N., & Brown, G. R. (2011). *Sense and Nonsense. Evolutionary perspectives on human behaviour* (2nd edition.). Oxford, UK: Oxford University Press.

Lamba, S., & Mace, R. (2011). Demography and ecology drive variation in cooperation across human populations. *Proceedings of the National Academy of Sciences*, *108*(35), 14426–14430. https://doi.org/10.1073/pnas.1105186108

Lewens, T. (2012). The Darwinian view of culture: Alex Mesoudi: Cultural evolution: how Darwinian theory can explain human culture and synthesize the social sciences. University of Chicago Press, 2011. *Biology & Philosophy*, *27*(5), 745–753. https://doi.org/10.1007/s10539-011-9306-y

Lewens, T. (2020). Cultural Evolution. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2020/entries/evolution-cultural/

Lozada, M., Ladio, A., & Weigandt, M. (2006). Cultural Transmission of Ethnobotanical Knowledge in a Rural Community of Northwestern Patagonia, Argentina. *Economic Botany*, *60*(4), 374–385. https://doi.org/10.1663/0013-0001(2006)60[374:CTOEKI]2.0.CO;2

Mace, R. (1996). Biased parental investment and reproductive success in Gabbra pastoralists. *Behavioral Ecology and Sociobiology*, *38*(2), 75–81. https://doi.org/10.1007/s002650050219

Mace, R. (1998). The co-evolution of human fertility and wealth inheritance strategies. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1367), 389–397. https://doi.org/10.1098/rstb.1998.0217

Mace, R. (2014). Human behavioral ecology and its evil twin. *Behavioral Ecology*, *25*(3), 443–449. https://doi.org/10.1093/beheco/aru069

Mace, R., & Pagel, M. (1994). The Comparative Method in Anthropology. *Current Anthropology*, *35*(5), 549–564.

Mace, R., Thomas, M. G., Wu, J., He, Q., Ji, T., & Tao, Y. (2018). Population structured by witchcraft beliefs. *Nature Human Behaviour*, *2*(1), 39–44. https://doi.org/10.1038/s41562-017-0271-6

Mace, R., & Zhang, H. (in press). Cross-cultural comparative methods for testing evolutionary hypotheses. In *The Oxford Handbook of Cultural Evolution*. Oxford, UK: Oxford University Press.

Mattison, S. M., Shenk, M. K., Thompson, M. E., Borgerhoff Mulder, M., & Fortunato, L. (2019). The evolution of female-biased kinship in humans and other mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1780), 20190007. https://doi.org/10.1098/rstb.2019.0007

Mesoudi, A. (2011). *Cultural Evolution. How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. Chicago, IL: The University of Chicago Press.

Micheletti, A. J. C. (2020). Modelling cultural selection on biological fitness to integrate social transmission and adaptive explanations for human behaviour. *Evolutionary Human Sciences*, *2*. https://doi.org/10.1017/ehs.2020.12

Micheletti, A. J. C., Brandl, E., & Mace, R. (2022). What is cultural evolution anyway? *Behavioral Ecology*, accepted.

Micheletti, A. J. C., Ge, E., Zhou, L., Chen, Y., Zhang, H., Du, J., & Mace, R. (2022). Religious celibacy brings kin-selected benefits. *OSF Preprints*.

Miguel, E. (2005). Poverty and Witch Killing. *The Review of Economic Studies*, *72*(4), 1153–1172. https://doi.org/10.1111/0034-6527.00365

Morgan, L. H. (1871). *Systems of consanguinity and affinity of the human family*. Washington, DC: Smithsonian Institution.

Murdock, G. P., Textor, R., Barry III, H., White, D. R., Gray, J. P., & Diwale, W. T. (1999). Ethnographic Atlas. *World Cultures*, *10*, 24–136.

Nettle, D., Gibson, M. A., Lawson, D. W., & Sear, R. (2013). Human behavioral ecology: current research and future prospects. *Behavioral Ecology*, *24*(5), 1031–1040. https://doi.org/10.1093/beheco/ars222

Norenzayan, A., Shariff, A. F., Gervais, W. M., Willard, A. K., McNamara, R. A., Slingerland, E., & Henrich, J. (2016). The cultural evolution of prosocial religions. *Behavioral and Brain Sciences*, *39*. https://doi.org/10.1017/S0140525X14001356

Oster, E. (2004). Witchcraft, Weather and Economic Growth in Renaissance Europe. *Journal of Economic Perspectives*, *18*(1), 215–228. https://doi.org/10.1257/089533004773563502

Pagel, M., & Mace, R. (2004). The cultural wealth of nations. *Nature*, *428*(6980), 275–278.

Passmore, S., & Jordan, F. M. (2020). No universals in the cultural evolution of kinship terminology. *Evolutionary Human Sciences*, *2*, e42. https://doi.org/10.1017/ehs.2020.41

Peacey, S., Campbell, O. L. K., & Mace, R. (under review). The sex of accused "witches" determined by competitive relationships.

Powers, S. T., Schaik, C. P., & Lehmann, L. (2021). Cooperation in large-scale human societies — What, if anything, makes it unique, and how did it evolve? *Evolutionary Anthropology: Issues, News, and Reviews*, *30*(4), 280–293. https://doi.org/10.1002/evan.21909

Qirko, H. (2002). The Institutional Maintenance of Celibacy. *Current Anthropology*, *43*(2), 321–329. https://doi.org/10.1086/339380

Ramsey, G., & De Block, A. (2017). Is Cultural Fitness Hopelessly Confused? *The British Journal for the Philosophy of Science*, *68*(2), 305–328. https://doi.org/10.1093/bjps/axv047

Rawlings, B., Dutra, N., Turner, C., & Flynn, E. (2019). Overimitation across development: the influence of individual and contextual factors. In *Conducting Research in Developmental Psychology*. New York, NY: Routledge.

Richerson, P. J., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., et al. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences*, *39*(e30), 1–68.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: how culture transformed human evolution*. Chicago, IL: The University of Chicago Press.

Ross, R. (2021). Is religious priming robust? (Talk at Cultural Evolution Society Conference 2021, Sapporo). Presented at the Cultural Evolution Society Conference 2021, Sapporo.

Salali, G. D., Chaudhary, N., Thompson, J., Grace, O. M., van der Burgt, X. M., Dyble, M., et al. (2016). Knowledge-Sharing Networks in Hunter-Gatherers and the Evolution of Cumulative Culture. *Current Biology*, *26*(18), 2516–2521. https://doi.org/10.1016/j.cub.2016.07.015

Schuppli, C., & Schaik, C. P. van. (2019). Animal cultures: how we've only seen the tip of the iceberg. *Evolutionary Human Sciences*, *1*. https://doi.org/10.1017/ehs.2019.1

Shariff, A. F., Willard, A. K., Andersen, T., & Norenzayan, A. (2015). Religious Priming: A Meta-Analysis With a Focus on Prosociality. *Personality and Social Psychology Review*, *20*(1), 27–48. https://doi.org/10.1177/1088868314568811

Sharpe, J. (1996). *Instruments of Darkness: Witchcraft in Early Modern England.* Philadelphia, PA: University of Pennsylvania Press.

Shenk, M. K., Begley, R. O., Nolin, D. A., & Swiatek, A. (2019). When does matriliny fail? The frequencies and causes of transitions to and from matriliny estimated from a de novo coding of a cross-cultural sample. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1780), 20190006. https://doi.org/10.1098/rstb.2019.0006

Shenk, M. K., & Mattison, S. M. (2011). The rebirth of kinship: evolutionary and quantitative approaches in the revitalization of a dying field. *Human Nature*, *22*(1–2), 1–15. https://doi.org/10.1007/s12110-011-9105-9

Singh, M. (2021). Magic, Explanations, and Evil: The Origins and Design of Witches and Sorcerers. *Current Anthropology*, *62*(1), 2–29. https://doi.org/10.1086/713111

Singh, M., & Henrich, J. (2020). Why do religious leaders observe costly prohibitions? Examining taboos on Mentawai shamans. *Evolutionary Human Sciences*, *2*. https://doi.org/10.1017/ehs.2020.32

Smith, D. (2020). Cultural group selection and human cooperation: a conceptual and empirical review. *Evolutionary Human Sciences*, *2*, e2. https://doi.org/10.1017/ehs.2020.2

Smith, E. A., & Winterhalder, B. (1992). *Evolutionary Ecology and Human Behavior*. New York, NY: Aldine de Gruyter.

Stout, D. (2013). Neuroscience of technology. In *Cultural evolution: society, technology, language, and religion*. Cambridge, MA: MIT Press.

Tanaka, M. M., Kendal, J. R., & Laland, K. N. (2009). From Traditional Medicine to Witchcraft: Why Medical Treatments Are Not Always Efficacious. *PLoS ONE*, *4*(4), e5192. https://doi.org/10.1371/journal.pone.0005192

Teixidor-Toneu, I., Jordan, F. M., & Hawkins, J. A. (2018). Comparative phylogenetic methods and the cultural evolution of medicinal plant use. *Nature Plants*, *4*(10), 754–761. https://doi.org/10.1038/s41477-018-0226-6

Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift für Tierpsychologie*, *20*(4), 410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Tooby, J., & Cosmides, L. (1990). The past explains the present. *Ethology and Sociobiology*, *11*(4–5), 375–424. https://doi.org/10.1016/0162-3095(90)90017-Z

Trivers, R. L. (1974). Parent-Offspring Conflict. *American Zoologist*, *14*(1), 249–264. https://doi.org/10.1093/icb/14.1.249

Tylor, E. B. (1903). *Primitive Culture. Researches into the development of mythology, philosophy, religion, language, art and custom.* (4th edition.). London: Murray.

van Elk, M., Matzke, D., Gronau, Q., Guang, M., Vandekerckhove, J., & Wagenmakers, E.-J. (2015). Meta-analyses are no substitute for registered replications: a skeptical perspective on religious priming. *Frontiers in Psychology*, *6*, 1365. https://doi.org/10.3389/fpsyg.2015.01365

Wertz, A. E. (2019). How Plants Shape the Mind. *Trends in Cognitive Sciences*, *23*(7), 528–531. https://doi.org/10.1016/j.tics.2019.04.009

Wertz, A. E., & Wynn, K. (2014). Thyme to touch: Infants possess strategies that protect them from dangers posed by plants. *Cognition*, *130*(1), 44–49. https://doi.org/10.1016/j.cognition.2013.09.002

Whiten, A. (2015). Experimental studies illuminate the cultural transmission of percussive technologies in *Homo* and *Pan*. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1682), 20140359. https://doi.org/10.1098/rstb.2014.0359

Wilson, E. O. (1975). *Sociobiology. The New Synthesis*. Cambridge, MA: Harvard University Press.

Winterhalder, B., & Smith, E. A. (2000). Analyzing adaptive strategies: Human behavioral ecology at twenty-five. *Evolutionary Anthropology: Issues, News, and Reviews*, *9*(2), 51–72.

Włodarczyk, A., Elsner, C., Schmitterer, A., & Wertz, A. E. (2018). Every rose has its thorn: Infants' responses to pointed shapes in naturalistic contexts. *Evolution and Human Behavior*, *39*(6), 583–593. https://doi.org/10.1016/j.evolhumbehav.2018.06.001

Wu, J.-J., Ji, T., He, Q.-Q., Du, J., & Mace, R. (2015). Cooperation is related to dispersal patterns in Sino-Tibetan populations. *Nature Communications*, *6*(1), 8693. https://doi.org/10.1038/ncomms9693

Zhang, H., Ji, T., Pagel, M., & Mace, R. (2020). Dated phylogeny suggests early Neolithic origin of Sino-Tibetan languages. *Scientific Reports*, *10*(1), 20792. https://doi.org/10.1038/s41598-020-77404-4

Zhang, H., & Mace, R. (2021). Cultural extinction in evolutionary perspective. *Evolutionary Human Sciences*, *3*, e30. https://doi.org/10.1017/ehs.2021.25

Zhou, L., Ge, E., Micheletti, A. J. C., Chen, Y., Du, J., & Mace, R. (under review). Monks relax sibling competition over parental resources in Tibetan populations.