EVOLUTIONARY EXPLANATIONS IN ECONOMICS

Vaios Koliofotis

EVOLUTIONARY EXPLANATIONS IN ECONOMICS

Vaios Koliofotis

Evolutionary Explanations in Economics

Vaios Koliofotis

ISBN 978-90-831893-5-2

2021

Print: Proefschriftenprinten.nl, The Netherlands

Front cover: A male *Maratus speciosus* peacock spider

Photo: Jurgen Otto

https://www.peacockspider.org/

EVOLUTIONARY EXPLANATIONS IN ECONOMICS

Evolutionaire verklaringen in de economische wetenschap

Thesis

to obtain the degree Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board

the public defence shall be held on

Thursday, 18 November at 10:30 hrs

by

Vaios Koliofotis
born in Athens, Greece

Erasmus University Rotterdam

Doctoral committee

Promotor:
Prof. dr. J.J. Vromen

Other members:
dr. C. Clarke
Prof.dr. LHJ Noordegraaf - Eelens
Prof.dr. G.M. Hodgson

Co-promotor: dr. C. Heilmann

The good seed that nature plants in us is so slight and so slippery that it cannot withstand the least harm from wrong nourishment.

Étienne de La Boétie (1548)

Table of Contents

Chapter 1: Introduction

1. Evolution and Economics

Economics and biology have been interacting throughout their histories. Thomas Malthus (1798) inspired Charles Darwin's theory of evolution by natural selection (Darwin, 1859). A population will undergo change if there is phenotypic variation, differential fitness between phenotypes and inheritance (Lewontin, 1970). In his paper with the characteristic title Why Is Economics Not an Evolutionary Science?, Thorstein Veblen (1898) urged economists to adopt a Darwinian approach. Influential economistssuch as Alfred Marshall (1920) recognized that economics could be benefit from biological reasoning. In an appendix of his classic book Principles of Economics, Marshall wrote that "… economics has no near kinship with any physical science. It is a branch of biology broadly interpreted".

More recently, Darwin's theory has influence to disciplines like psychology (e.g. Cosmides and Tooby, 1992) and economics (e.g. Vromen, 1995) and the three core Darwinian principles are important in philosophical discussions of evolutionary change. Within economics, particular lines of research explicitly draw upon concepts from evolutionary theory (e.g. Hodgson, 2013). Bioeconomics (e.g. Landa and Ghiselin, 1999), Evolutionary Economics (e.g. Nelson and Winter, 1982) and Evolutionary Behavioral Economics (e.g. Burnham et al., 2016) are interdisciplinary sub-fields that incorporate insights from biology into economic theory. They emerged after the 1970s and the pioneering work of researchers such as Gary Becker (1976), Jack Hirshleifer (1977) and Nelson and Winter (1982).

This thesis discusses key concepts of evolutionary theory and their application to economic issues. It adopts the perspective that economic choices and their outcomes are influenced by individual characteristics or traits. Since some human traits were shaped by evolutionary processes, there is substantial potential for the use of evolutionary biology in economic analysis. For example, human preferences related to nutrition, mating, maternal care, sociality and novelty play a central role in economic decision-making and by understanding the evolutionary foundation of these preferences, it is possible to study choices in and outside markets.

On the intersection of economics and biology, evolutionary theory has been used to examine particular traits, their genetic (molecular) basis and the role of evolutionary dynamics to economic growth. The subject matter of this thesis needs to be distinguished from what is called 'Evolutionary Economics'. The emergence of evolutionary economics can be traced to Nelson and Winter (1982), whose seminal

work inspired a vast literature on firm innovation and economic growth. The subject matter of this thesis differs from evolutionary economics in that I focus on human traits rather biological concepts applied to the dynamics of firms and institutions. Moreover, the topics discussed in the chapters that follow have to be distinguished from important empirical findings in genoeconomic studies on heritable economic traits (e.g. Benjamin et al., 2012). What suffices for the arguments presented in this thesis is the uncontested assumption that a proportion of the variation in traits relevant to economists can be attributed to genetic variation.

The recent growth in our knowledge of evolutionary processes is typically organized around key distinctions that can be easily applied to different topics that are relevant to economics. In particular, philosophers and scientists have sought to draw distinctions among different kinds of explanations. This dissertation is arranged around two particular distinctions: the distinction between ultimate and proximate explanations (Chapters 2-4), and the distinction between causal and non-causal explanations (Chapter 5).

In what follows, I focus on Mayr's ultimate-proximate distinction and I provide a general account of what it involves. This will clarify how formal evolutionary theory makes claims regarding particular traits. In chapter 5 I will add another distinction, concerning causal and non-causal explanations. In articulating these distinctions in the chapters that follow, this thesis contributes to the resurgence of interest in using evolution theory for the study of economics traits, similar to the way that it is applied in the biological sciences.

2. The Ultimate – Proximate Distinction Reloaded

A 1961 article by Ernst Mayr titled "Cause and effect in biology: Kinds of causes, predictability, and teleology are viewed by a practicing biologist" has become a classic for summarizing how evolutionary theory provides a framework for studying human traits. Mayr introduced the ultimate-proximate distinction that researchers often apply to dispel misunderstandings of evolutionary theory in biology and the social sciences.

According to Vromen (2017), one can find in Mayr's paper at least three aspects to the ultimate – proximate distinction: First, Mayr drew a distinction between evolutionary causes (i.e. natural selection) and proximate trait-generating causes (e.g. psychological, neurological mechanisms) inside organisms. Second, there is also a distinction between distant and nearby parts in the causal chain that leads to a trait. Finally, there is a distinction between different research questions: ultimate explanations typically answer why-questions while proximate explanations typically answer how-questions.

It is useful to explicate the different sorts of research questions that can be asked about traits and the type of answers they might call for. The ultimate question is the following: Why did a particular behavior evolve compared to other behaviors that could have evolved in the lineage? One approach to answering this question is to invoke evolutionary processes that are temporally distant in the causal chain that leads up to a particular trait. Evolutionary researchers are also concerned with behavioral responses to current environmental conditions and their fitness outcomes. For example, when a social behavior like cooperation is explained in terms of the dynamic of natural selection and how, on average, it results to positive fitness consequences, researchers provide an ultimate explanation. Not all behaviors have evolved by natural selection. Some are the result of random chance events (i.e. drift) or are byproducts that do not have any beneficial fitness effects.

The dominant account of ultimate explanation is the causal one and most evolutionary researchers assume that their evolutionary models identify the causes of human and animal behavior (e.g. Okasha and Martens 2016). In philosophy of science, however, a growing body of literature has challenged the predominance of causal explanation by investigating putative cases of non-causal explanations (e.g. Lange 2013; Reutlinger 2017). The main contention in chapter 5 of this thesis is that there are evolutionary models that do not explain in virtue of citing causes. Instead, they provide a non-causal explanation in accord to recent accounts suggested in the literature. For example, evolutionary models might provide an ultimate explanation by constrain or a non-causal counterfactual ultimate explanation.

The proximate question is how doesthe interaction of the external environment and the internal mechanism produce behavior?Proximate explanations are not concerned with why a particular behavior evolved but with how it works. In proximate explanations, behavior is typically understood in mechanistic terms. The idea here is that a proximate mechanism results to a behavioral expression in a particular environment. For example, when a social behavior such as cooperation is explained in terms of rational decision-making, behavior-generating social preferences or social learning biases, researchers provide a proximate explanation.

The need to distinguish between proximate and ultimate questions, causes and explanations is one of the key insights of employing an evolutionary account. It is generally accepted that there is no contrast between proximate explanations that invoke behavior-generating mechanisms and ultimate explanations based on natural selection. Ultimate and proximate explanations complement each other and both are required for a complete understanding of a particular trait.

The proximate-ultimate distinction has been the cornerstone of evolutionary research. However, in the chapters that follow it is argued that a strong commitment to this distinction is not always useful for evolution-based behavioral economics. Examining only economic behavior can draw attention away from general patterns of proximate mechanisms that evolutionary theory can uncover and explain. Standard evolutionary models focus on genetic variation and behavioral outcomes. Proximate mechanisms are not completely ignored, but they are studied after

evolutionary models are built and almost never integrated to an evolutionary account of human behavior. One notable exception discussed in this thesis is social learning mechanisms that figure is ultimate explanations of behavior provided bygene-culture co-evolution theory. In contrast, economics puts much emphasis on proximate mechanisms using models of decision-making and psychology. Focusing on preferences and beliefs is irrelevant to most theoretical biologists. In economics, the concrete structure and theoretical role of preferences and decision-making are considered topics worthy of detailed analysis.

Mayr stressed that the proximate how question and the ultimate why question need to be addressed for all evolutionary products. It is possible, however, to raise a different research question when it comes to explaining human evolution. What is important to observe is that behavior is brought about by psychological mechanisms, at an intermediate level between genes and behavior. Hence, we can go beyond Mayr's distinction and pose the following question: Why did a particular proximate mechanism evolve rather than some other that could have produced a set of human behaviors? To answer this question, evolutionary researchers can assume that behavioral patterns expressed in a specific environment are not directly under selection but are the product of evolved psychological mechanisms. To put it in another way, in the ultimate explanations advanced in this thesis, natural selection produces behavior indirectly through proximate causes.

This mechanism selection question suggests an alternative line of theoretical research. It requires that we place at the centre of analysis proximate mechanisms, rather than the behavioral outcomes they produce. If we already have a good understanding of decision-making or psychological mechanisms underlying behavior, evolutionary theorizing might help us explain why they have evolved. The underlying idea here is that if proximate mechanisms are treated as the outcome of evolutionary processes, evolutionary theorizing may generate interesting hypotheses about the sorts of proximate mechanisms that might have evolved. While current evolutionary modeling treats psychological mechanisms as a constraint on the behavioral repertoire, in this thesis I contend that the mechanisms underlying human behaviors can be viewed not as a constraint but as an evolved trait. In the chapters that follow I want to demonstrate that this perspective can turn out to be fruitful and can lead to new insights in evolution-based economics.

3. Putting Philosophy to Work

The mechanism selection question suggests a new way of thinking about the evolution of some human traits. In what follows I discuss the structure of the thesis and how the mechanism selection question fits into current debates in biology, economics and philosophy.

Before I turn to a detailed explication of each chapter, let me first provide a brief summary followed by an explanation of the logic of the organization of this dissertation. This thesis concerns research topics in which economics, biology and philosophy overlap. It contains four essays in which individual economic choices and outcomes are related to evolutionary processes such as natural selection. In the next chapter, I introduce an evolutionary method while chapters 3 and 4 apply the findings on methodology from chapter 2 to analyze two traits that are of interest to economics: conspicuous consumption and social emotions. The final essay focuses on causes, correlates and consequences in ultimate explanations of social behavior (chapter 5). Each chapter provides a more detailed introduction and a literature review of topics discussed in this thesis.

In more detail, chapter 2 presents a methodological framework that can be applied to evolutionary research in economics. Based on the difference between two kinds of formal theorizing in biology and economics, the fundamental point of the first essay is to construct, explore, and defend a method of examining the evolution of human proximate mechanisms. I argue that a distinction between two kinds of evolutionary theorizing highlights differences in methodology. Human behavioral ecology and evolutionary game theory focus on behaviors and fitness outcomes. In contrast to models in formal biology, economists have developed the indirect evolutionary approach which is concerned with articulating the concrete mechanisms and the evolutionary function, through time, of preferences and behavior.

Based on the notion of the "Adaptively Relevant Environment" and fitness maximization, in chapter 2 I take a critical view of a method proposed by Burnham (2013). I suggest that many models in evolutionary game theory deliver incomplete explanations, as they do not examine specific historical (ethnographic, archeological etc) data for human populations and they ignore that genes have their effects to behaviour through proximate mechanisms. Hence, I emphasize two points: First, the significance of studying past environments and second, the potential benefits of shifting the focus of theoretical research to the study of the evolution of proximate mechanisms governing human psychology and decision-making. The evolutionary methodology developed in chapter 2 is used in next two chapters of the dissertation to examine particular traits.

In chapter 3, I focus on conspicuous consumption. A number of papers draw upon ideas from sexual selection and costly signaling theory to argue that conspicuous consumption is a mating strategy evolved by sexual selection to provide reproductive benefits to men. Applying tools from chapter 2, I argue that evolutionary researchers attempt to examine the evolutionary underpinnings of conspicuous consumption without analyzing in detail human evolved psychology and the available evidence about the ancestral environment. After I outline the main criteria for arguing that conspicuous consumption is the outcome of sexual selection, I argue that on closer examination this explanation does not necessarily chime with our knowledge of ancestral environments. Moreover, I apply insights

from cultural evolution theory and present evolved proximate mechanisms that might be involved in consumption practices. In particular, conspicuous consumption can be understood as a pattern of behavior marked by a set of social learning mechanisms. By explicating particular proximate learning mechanisms I address the possibility of evolutionary research that focuses on human psychology and evolutionary process.

In chapter 4, I contribute to recent work on the evolution of human cooperation. Robert Frank's work is the most popular attempt in economics to explain social emotions, non-verbal expressions and social behaviors such as cooperation and altruism (e.g. Frank, 1988). Following this line of work, I focus on the criticisms raised against the controversial green-beard explanation of social behaviours. Based on conceptual and theoretical resources found in Frank's work, I propose an alternative red-beard explanation of human sociality. In line with chapter 2, this explanation focuses on the emergence and evolution of human emotions, a proximate cause, rather than patterns of behavior. I demonstrate that proximate mechanisms can be a key to the stability of cooperation. The main idea in this chapter is that social emotions are a common cause of a social behavior and a phenotypic marker (for example, a non-verbal facial expression) and therefore cooperative behavior cannot be suppressed without also changing the marker.

In chapter 5, **I** examine the debates that have erupted in the last decade over the status of inclusive fitness theory and Hamilton's Rule (e.g. Hamilton 1964a, 1964b, Nowak et al., 2010; Abbot et al., 2011). Inclusive fitness theory is one of the most important theoretical extensions to Darwinian Theory and the regression version of Hamilton's Rule facilitates contemporary mathematical research on social evolution. Together with the ultimate-proximate distinction, they have been applied to examine whether a series of evolutionary models developed by economists (e.g. Gintis, 2000) identify a novel pathway to the evolution of social traits (e.g. Vromen, 2012).

While the proximate-ultimate distinction refers to the causal structure of explanations provided by evolutionary theory, chapter 5 focuses on the distinction between causal and non-causal explanation in mathematical models. The aim is to examine the explanatory power of the general version of Hamilton's Rule which has been a point of contention within the literature (Gardner et al., 2011; Nowak et al, 2017). After introducing the controversy over the rule's explanatoriness, this chapter discusses Birch's suggestion that Hamilton's Rule explains qua being an organizing framework (Birch, 2017). Together with Philippe Verreault-Julienwe argue that Hamilton's Rule does not explain by virtue of citing causes and propose a different view of its explanatoriness. Using resources from a recent debate in philosophy of science on non-causal explanations, we suggest that the regression version of Hamilton's Rule either provides a non-causal counterfactual explanation (Woodward, 2018) or an 'explanation by constraint' (Lange 2017). Hence, the analysis in this chapter reinforces the case for the considerable value of the regression-based version of the rule in explaining the evolution of social traits.

4. Relevance

This thesis sets out the challenging task to expound the application of evolutionary reasoning to economics. It provides an in-depth analysis to a cluster of arguments and topics of research that are likely to be relevant to those working at the intersection between philosophy, economics and evolutionary biology and are interested in new ways of thinking about human traits. Moreover, it provides useful background knowledge required to follow recent debates in evolutionary biology and economics. In the final section of the introduction I discuss the implications and relevance of the findings of this thesis. I expound the contribution of this thesis to different lines of future current research that can is be followed by economists and philosophers interested in evolutionary theory.

The topics analyzed in the previous sections are closely related to recent attempts in Behavioral Ecology to discuss assumptions implicit in evolutionary models. Evolutionary scientists often employ what Grafen calls the phenotypic gambit, the assumption that traits can be studied without understanding the underlying genetic and psychological mechanisms (Grafen, 1984). This is a controversial assumption and there are cases where it may be wrong (e.g. Fawcett et al, 2012). The genetic structure and the way that evolution has shaped human psychology and decision-making may have an important influence to human and animal behavior.

This thesis can be also seen as a positive contribution to evolutionary research promoting an Extended Evolutionary Synthesis (EES) over the past decade (e.g. Laland et al. 2015). The synthesis between the Darwinian theory of evolution and Mendelian genetics of inheritance is based on the distinction between proximate and ultimate explanation of behavior. The proximate-ultimate distinction has been used to establish what can be considered a satisfactory ultimate explanation. In this thesis, however, proximate mechanisms that cause individuals to have particular behavior do appear as an explanandum in ultimate explanations. This is also a commitment of the EES position; proximate mechanisms both shape and respond to selection.

A final note is in order. This thesis attempts to be balanced in terms of the attention given to philosophy, biology and economics. I have chosen my topics to achieve two goals. First, I want to show that philosophy can be made applicable and relevant to analyze and clarify biological and economic practice, both empirical and theoretical. It is not always clear whether tools from philosophy can be made relevant to debates in fields like biology and economics that contribute to our understanding of human traits. The second goal is to analyze evolutionary processes from a different perspective and provide an in-depth analysis of particular traits relevant to economists. The aim is to bring modern evolutionary theory to the attention of the

economics community and contribute to the growth of evolutionary theorizing in economics.

References

Abbot, P., J. Abe, J. Alcock, S. Alizon, J. A. C. Alpedrinha, Ma. Andersson, et al. (2011). Inclusive Fitness Theory and Eusociality. Nature 471 (7339): E1–4

Allen, B., M. A. Nowak, and E. O. Wilson. (2013). Limitations of Inclusive Fitness.Proceedings of the National Academy of Sciences 110 (50): 20135–39

Becker, G. (1976). Altruism, Egoism and Genetic Fitness: Economics and Sociobiology. Journal of Economic Literature 14(3), 817–826

Benjamin, D. J., D. Cesarini, C. F. Chabris, E. L. Glaeser, D. I. Laibson, et al. (2012). The Promises and Pitfalls of Genoeconomics. Annual Review of Economics 4(1), 627–662.

Birch, J. (2017).The Philosophy of Social Evolution. Oxford: Oxford University Press.

Burnham, T. C. (2013).Toward a neo-Darwinian synthesis of neoclassical and behavioral economics.Journal of Economic Behavior & Organization, 90, S113–S127.

Burnham, T. C., Stephen, E. G., Lea, S. E., Bell, A., Gintis, H., Glimcher, P. W., et al. (2016). Evolutionary behavioral economics. In D. S. Wilson & A. Kirman (Eds.), Complexity and evolution: toward a new synthesis for economics. Strungmann Forum Report 19. Cambridge, MA: The MIT Press

Cordes, C. (2019). The promises of a naturalistic approach: how cultural evolution theory can inform (evolutionary) economics. Journal of Evolutionary Economics, 29(4), 1241–1262.

Cosmides, L. and J. Tooby (1992) Cognitive Adaptations for Social Exchange, in: Barkow, J., Cosmides, L., Tooby, J. (eds.), The Adapted Mind: Evolutionary Psychology and the Generation of Culture. New York: Oxford University Press, pp. 163–228.

Darwin, C. (1871). The descent of man and selection in relation to sex. J. Murray, London

Fawcett, T. W., Hamblin, S., &Giraldeau, L.-A. . (2012). Exposing the behavioral gambit: the evolution of learning and decision rules. Behavioral Ecology, 24(1), 2–11.

Frank, R. H. (1988). Passions Within Reason: The Strategic Role of the Emotions. New York: Norton

Gardner, A., S. A. West, and G. Wild. (2011). The Genetical Theory of Kin Selection.Journal of Evolutionary Biology 24 (5): 1020–43.

Gintis, H. (2000). Strong Reciprocity and Human Sociality. Journal of Theoretical Biology 206 (2): 169–79.

Grafen, A. (1984). Natural selection, kin selection and group selection. In J. R. Krebs, & N. B. Davies (Eds.), Behavioural ecology: An evolutionary approach, pp. 62–84. Oxford: Blackwell Scientific

Hamilton, W.D. (1964a) The Genetical Evolution of Social Behaviour I. Journal of Theoretical Biology 7, 1–16.

Hamilton, W.D. (1964b) The Genetical Evolution of Social Behaviour II. Journal of Theoretical Biology 7, 17–52.

Hirshleifer, J. (1977). Economics from a Biological Viewpoint. The Journal of Law and Economics, 20(1), 1–52.

Hodgson, G. (2013). From Pleasure Machines to Moral Communities: An Evolutionary Economics without Homo economicus. University of Chicago Press, Chicago

Landa, J. T. and M. T. Ghiselin. (1999). The emerging discipline of bioeconomics: aims and scope of the Journal of Bioeconomics. Journal of Bioeconomics 1: 5–12.

Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., and Odling-Smee, J. (2015). The extended evolutionary synthesis: its structure, assumptions and predictions. Proceedings of the Royal Society B: Biological Sciences, 282(1813), 20151019.

Lange, M. (2013). What Makes a Scientific Explanation Distinctively Mathematical? The British Journal for the Philosophy of Science 64 (3): 485–511.

Lange, M. (2017). Because Without Cause: Non-Causal Explanations in Science and Mathematics. New York: Oxford University Press.

Lewontin, R. C. (1970). The Units of Selection." Annual Review of Ecology and Systematics 1: 1-18.

Malthus, T. (1798). An Essay on the Principle of Population. London: Printed For J Johnson

Marshall, A. (1920). Principles of Economics. London: Macmillan.

Mayr, E. (1961). Cause and Effect in Biology: Kinds of causes, predictability, and teleology are viewed by a practicing biologist. Science 134(3489), 1501–1506

Nelson, R. R., and Winter, S. G. (1982). An evolutionary theory of economic change. Cambridge, Mass. The Belknap Press Of Harvard Univ. Press.

Nowak, M., A. McAvoy, B. Allen, and E. O. Wilson. (2017). The General Form of Hamilton's Rule Makes No Predictions and Cannot Be Tested Empirically. Proceedings of the National Academy of Sciences 114 (22): 5665–70.

Reutlinger, A. (2017). Explanation Beyond Causation? New Directions in the Philosophy of Scientific Explanation." Philosophy Compass 12 (2): 1–11.

Okasha, S. and J. Martens. (2016). The Causal Meaning of Hamilton's Rule.Royal Society Open Science 3 (3): 160037.

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

Vromen, J. (1995). Economic evolution: an enquiry into the foundations of new institutional economics. Routledge, London

Vromen, J. (2012). "Human Cooperation and Reciprocity." In Evolution and Rationality: Decisions, Co-Operation and Strategic Behaviour, edited by Samir Okasha and Ken Binmore, 158–84. Cambridge: Cambridge University Press.

Vromen, J. (2017). Ultimate and proximate explanations of strong reciprocity.History and Philosophy of the Life Sciences, 39(3).

Woodward, J. (2018). "Some Varieties of Non-Causal Explanation." In Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations,edited by Alexander Reutlinger and JuhaSaatsi, 117–38. Oxford: Oxford University Press

Chapter 2: Applying Evolutionary Methods in Economics: Progress or Pitfall?[1]

1. Introduction

Over the past decades there is a growing interest in building bridges between evolutionary biology and economics. Widely cited studies published in leading economic journals apply evolutionary arguments and theories to explain and discover features of human psychology and cognition that have a role in economic behavior (for recent reviews, see Collins et al., 2016, Burnham et al., 2016).

Many economists hold that the insights of evolutionary theory might be relevant for economics in that they can help improve our understanding of human preferences and behavior. Looking back into the writings of Jack Hirshleifer, a pioneer economist that examined what evolutionary biology can contribute to economics, "[t]he programmatic contention here is that such preference patterns, despite seemingly arbitrary elements, have survived because they mainly are adaptive to environmental conditions" (Hirshleifer, 1977: p.18). In the economic literature human preferences are taken as given. For Jack Hirshleifer, evolutionary theory could help economics with identifying what preferences we have and explaining why they survived selection processes (Hirshleifer, 1977: p.17).

It is still unclear, however, what method can facilitate economists in formulating evolutionary hypotheses about human preferences and behavior. Terence Burnham presents a promising evolutionary method that aspires to reconcile different sub-fields within economics (Burnham, 2013). In this chapter, I set aside prospects for a broad integration within economics or among the social sciences. My relatively modest aim is to clarify and refine Burnham's proposal. It is also not my intention to present and critically assess alternative methods offered by evolutionary psychologists[2]. Instead, I examine Burnham's method on its own terms and consider whether it can be utilized to inspire productive work at the interface between economics and biology.

In Burnham's view, it is possible to formulate novel hypotheses and predictions about human preferences, on the basis of what patterns of behavior would have been adaptive in the Adaptively Relevant Environment (henceforth ARE). After

---

[1]This chapter is published as Koliofotis, V. Applying evolutionary methods in economics: progress or pitfall?. *J Bioecon* (2020). https://doi.org/10.1007/s10818-020-09300-w

[2] For an insightful discussion of these methods, see Griffiths (1996).

clarifying two crucial theoretical terms that figure in Burnham's method, the ARE and maximizing behavior, I suggest that identifying the ARE requires economists to describe a particular trait and build or pick an evolutionary model that identifies a set of selection pressures against which it would confer an evolutionary advantage. With this in mind, I further argue that evolutionary models of adaptive behavior do not routinely entail particular insights about human psychology. The problem this raises for Burnham's method is that several scientifically plausible hypotheses about human preferences will be compatible with such evolutionary models. Finally, based on the theoretical resources provided by the indirect evolutionary approach (henceforth IEA), I argue that Burnham's method can be more reliable if economists apply or build evolutionary models of their own that target particular features of human psychology and cognition. Such models may reduce the number of plausible hypotheses to allow for rigorous scientific testing in laboratory or field experiments.

Based on these arguments, the revised method proposed in this chapter is the following:

i. Identify preferences and overt behaviors related to them.
ii. Apply or build an evolutionary model and use information about our ancestral past to determine the ARE of these traits;
iii. Predict the structure of preferences and how they operate in the present environment;
iv. Test these predictions in the laboratory and the field.

This chapter is structured as follows. In the next section I introduce Burnham's method, focusing on concepts that are more relevant for its application in theoretical and empirical research. Moreover, I set forth what I consider to be the main problems with his proposal. Following this, in section 3, I examine what kind of evolutionary models can be applied to shed light into our psychological traits. Section 4 explores in more depth whether it is reasonable to assume that decisions in the ancestral past were based on utility maximization. Section 5 concludes.

2. Burnham's Method and the Adaptively Relevant Environment

This section begins by presenting Burnham's proposal and focuses on two key concepts that are in need of disambiguation: The Adaptively Relevant Environment and maximizing behavior. Next I examine whether it is possible to search for and subsequently frame hypotheses and predictions about human preferences, using only evolutionary models that explain patterns of behavior.

Burnham's research strategy involves the following four steps:

i. Provide a model of the Adaptively Relevant Environment (ARE);

ii. Solve for maximizing behavior in the ARE and predict the structure of preferences;
iii. Predict how these evolved preferences in the ARE operate in the present environment;
iv. Test these predictions and refine.

According to Burnham, economists begin their research by examining the human ancestral environment, trying to find therein insight into what sorts of behaviors could have evolved by selection. In particular, economists identify human behavior that was maximized in the ARE. For Burnham, this is sufficient to infer hypotheses about our psychology and cognition and predict the structure of preferences that produce the relevant pattern of behavior. After examining how these preferences operate in the current environment, a test of these predictions is performed in the laboratory or the field.

Burnham's method is related to the following evolutionary hypothesis. Human behavior is the product of psychological traits that evolved in the ancestral environment. We, however, live in an environment that is very different from the one in which these traits evolved. Back in the Pleistocene savannah, for example, there were no fast food restaurants, contraceptives, antibiotics or automobiles. In a novel environment, our psychological traits often result in dysfunctional behaviors. If significant changes in the environment actually occurred, one might expect that many human behaviors would be detrimental to our survival or reproduction. Kanazawa (2004) calls this the Savannah Principle, Hagen and Hammerstein (2006) the mismatch hypothesis, Richerson and Boyd (2005) the adaptive-lag hypothesis and Burnham and Johnson (2005) the evolutionary legacy hypothesis (henceforth ELH).

The standard example is our preference to eat food containing sugar, fat and salt. This preference evolved in an environment where such substances were important to health, though they we difficult to obtain. In current environments however it is easy to buy and store food with sugar and fat. Our preference to eat such food evolved in the ancestral environment and is still present today, with obvious negative effects to our health.

At first sight, Burnham's proposal is neat with clear-cut steps that are easy to follow. But on closer look, it turns out to that key concepts are ambiguous and in need of clarification.

An important concept in Burnham's method is the ARE (Irons, 1998). In evolutionary research, many traits or characters (e.g. psychological, cognitive, morphological, behavioral) evolved in the context of the selection pressures that operated in the ancestral past. Selection pressures typically refer to features of the environment that exert an influence on reproductive success, driving the evolution of traits. They can take many forms, such as climatic conditions, habitat, mates,

migration, geographical area, availability and quality of food, energy resources, predators, diseases, population size as well as other conspecifics.

The ARE describes the main selection pressures thought to have led to the evolution of a particular trait throughout its evolutionary history[3] . However, evolutionary researchers should not equate the ARE with the Pleistocene, a period that our ancestors spent as hunter-gatherers in the African savannah, prior to the emergence of agriculture about 10,000 years ago. Some features of the African savannah may have had little to do with the evolution of particular traits. The ARE is also not necessarily tied to a particular epoch like the Pleistocene or the Holocene. Some traits have a much longer evolutionary history or have evolved much more recently in the last 10000 years. Instead, the ARE describes key features of different environments that influenced the evolution of a particular trait.

How can economists reconstruct the ARE? Burnham is silent on this issue. As I see it, evolutionary theory can help us track the selection pressures that might have led to the evolution of particular traits. An economist interested in evolutionary research has to describe in some detail a particular trait and pick or build an evolutionary model that identifies a set of selection pressures against which that particular trait could have evolved. If the selection pressures identified by the evolutionary model are supported by empirical evidence, evolutionary researchers can infer that they are part of the ARE.

This implies that one has to add a step in Burnham's method. Knowledge of a trait comes first and the application of an evolutionary theory to determine the ARE follows. Evolutionary theory, as any other theory, typically aims to explain traits that are known to be present in contemporary or past populations. Without some initial information about the target trait, it is not possible to describe it in some detail and build an evolutionary model that identifies its ARE.

The second key concept in Burnham's method is "maximizing behavior". The term is however ambiguous since Burnham does not clarify the maximand: Is it fitness, utility, income, well being or happiness? In what follows I will make an attempt to shed some light on this term by taking into consideration how "maximizing behavior" is typically understood in evolutionary biology.

Evolutionary scientists often suggest that a type of behavior is adaptive if it maximizes fitness, i.e. reproductive success or a currency conducive to survival and reproduction. A generalization of Darwinian fitness in case of social behaviors is

---

[3] Evolutionary researchers often apply the distinct concept of environment of evolutionary adaptedness (EEA for short), the environment "that man inhabited for two million years until changes of the past few thousand years led to the extraordinary variety of habitats he occupies today" (Bowlby, 1969: p. 58). In contrast to the EAA that describes a uniform background to which our traits adapted in the ancestral past, ARE is trait-specific.

inclusive fitness, the sum of the direct fitness gained through the influence of an individual's behavior on her production of offspring and the indirect fitness derived from contributing to the production of offspring of genetically related individuals.

A recent theoretical controversy revolves around the generality of inclusive fitness maximization and questions whether it can be used as a tool for understanding the evolution of human and animal traits. According to Alan Grafen's work, evolutionary researchers can expect fitness maximizing behavior to evolve only under particular conditions (Grafen, 2005; 2014). An important theoretical conclusion is that inclusive fitness maximization does not apply to cases of frequency-dependent selection and non-additive fitness effects[4]. Whether Grafen's results can be generalized and extended to the more realistic nonadditive case is an open theoretical question and has not yet been settled.

It is beyond the scope of this chapter to advance particular theoretical arguments for or against the principle of inclusive fitness maximization (e.g. Lehmann, Alger and Weibull, 2015). Given that this principle is controversial and its generalization has not yet been achieved, there is no reason to incorporate it in an evolutionary methodology that can be applied by economists and other social scientists. Moreover, such a principle is not necessary to the study of the evolution of human behavior. Instead of assuming that an organism's behavior maximizes inclusive fitness, evolutionary researchers can apply standard evolutionary models that examine whether particular traits were favored or disfavored by natural selection.

Let us stop for a moment and see where we have come. Burnham presents a methodology whereby economists begin their research with the ARE and identify therein maximizing (adaptive) behavior. A detailed examination of two key concepts, the ARE and fitness maximization, suggests the following refinement of Burnham's method. To determine the ARE economist have to start their investigation with a particular trait. And instead of taking for granted the controversial principle of (inclusive) fitness maximization, they can apply an evolutionary model that explains why this trait evolved in the ancestral past. After identifying the ARE of that trait, one can infer hypotheses about the presence of particular psychological traits. With these points in mind, one can identify two problems with Burnham's method that call for another refinement.

First, according to the ARE concept, economists cannot simply assert that we live in a novel environment, distinct from the Pleistocene environment, as Burnham and other evolutionary researchers have done in a variety of cases. Economists might anticipate that modern industrial societies are different from Pleistocene hunter-

---

[4] Alan Grafen, one of the proponents of the optimization programme in evolutionary biology, writes that "the assumption of additivity is made throughout this chapter, but is not in general a realistic assumption. In many applications, non-additivity is an important part of the problem" (Grafen, 2006: p. 543).

gatherer societies. However, it is entirely possible that some key features of ancestral environments that led to the evolution of a specific trait have remained roughly the same. The ARE is always identified for a particular trait and different traits often evolved against a different ARE. Without a detailed analysis of the trait of interest and the relevant features of past and present environments, economists cannot know to what extent the ARE differs from relevant features of the current environment.

Second, there is a problem of inference in the second step of Burnham's method, the move from models that target behavior to hypotheses about human preferences. What is important to emphasize is that when evolutionary theorists build models that assume fitness-maximization, their target of analysis is behavior, not human psychological or cognitive traits. Evolutionary game theory and evolutionary optimality models for example do not suggest that evolutionary processes select psychological traits that generate fitness-maximizing behavior. Instead, they remain agnostic about the specific neural, cognitive or psychological traits that guide animal and human behaviors. What these models explain is why patterns of behaviors are selected in the ancestral past. To the extent that multiple underlying psychological traits result in the same fitness adaptive behavior, Burnham's method will not be a powerful guide in telling which hypotheses about human psychology are more plausible. To the economist applying this method, many hypotheses about preferences appear to be just as good as any other.

The more general problem is that economists cannot simply assume there is only one psychological trait associated with adaptive behavior. Such a one-to-one mapping is very unlikely. It is plausible, I think, to accept that psychological traits are typically involved in behavior and it is through behavior that these traits confer to an organism a fitness advantage. Any given adaptive behavior however can be the result of a variety of psychological traits and each of them might have a different evolutionary history. The implication is that economists cannot lose sight on the numerous relations between adaptive behavior and psychological traits. And unless there are reasons independent of evolutionary models to expect that there are not many different psychological traits that can potentially support the same adaptive behavior, Burnham's method would not be a reliable guide in formulating hypotheses about human psychology [5].

---

[5] Although Burnham does not present additional resources that can help researchers to constrain hypotheses about psychological traits, using knowledge from cognitive psychology is a reasonable suggestion. This, however, raises the following problem: Cognitive science is not unified at the theoretical level. There are multiple theories (e.g. modularity hypothesis, heuristics and biases, folk-psychology, dual-process theory etc) and no consensus view has emerged as to which of these theories is more plausible. I discuss this issue in more detail in section 4.

Before I tackle this problem in the next section, it is useful to give empirical content to Burnham's proposal. He argues that his methodology provides the following novel hypothesis: High levels of altruistic behavior observed in economic experiments (and contemporary societies) arise from human psychological and cognitive traits that evolved in an ancestral environment of repeated interactions, and reputation was constantly at stake.

To arrive to this hypothesis, evolutionary researchers start with a general description of the environment faced by our Pleistocene hunter-gatherers: They lived in small groups and repeated interactions with other group members may have been an important feature of their lifestyle. In this sort of environment, it would have been adaptive for our ancestors to cooperate and protect their reputation. Burnham and other evolutionary researchers suggest that altruistic behavior in one-shot and anonymous interaction in contemporary environments is the result of human psychology fine-tuned to operate in repeated and public interactions. Evolutionary researchers conduct a series of experiments designed to test predictions derived from this hypothesis[6].

What I emphasized so far is that evolutionary researchers have to apply an evolutionary model to determine the ARE for a particular trait. Instead, evolutionary researchers provide a very general description of the hunter-gatherer lifestyle, without a detailed analysis of the actual selection pressures that led to the evolution of social behavior (e.g., Burnham and Johnson 2005; Haley and Fessler, 2005). In particular, it is not clear which evolutionary theory and therefore which adaptive behaviors and selection pressures these researchers consider to derive their hypothesis from.

For some researchers, direct reciprocity theory (Trivers, 1971) is assumed to be largely correct and cooperation among non-kin evolved in repeated interactions (Burnham and Johnson, 2005: p.130). In evolutionary biology, however, reputational effects are analyzed by indirect reciprocity theory (e.g. Nowak and Sigmund 2005). Both indirect and direct reciprocity models can explain the evolution of human cooperation. They demonstrate that cooperation evolves because it ultimately results in a net fitness benefit. The behaviors however examined by these theories are very different. The evolution of cooperation by direct reciprocity is often related to the success of "Tit-for-Tat" (TFT) in computer tournaments (Axelrod, 1984). This type of behavior cooperates on the first round

---

[6] Empirical support for this hypothesis is mixed. Many experiments demonstrate that reputation-related cues such as displaying images of eyes promote cooperative behavior (e.g. Haley and Fessler 2005; Burnham and Hare, 2007). Other experimental studies suggest that being observed by such cues do not affect experimental behavior (e.g. Fehr and Schneider, 2010). There is currently an ongoing debate about interpretation of experimental results (a short review can be found in Nettle et al. 2013).

and thereafter it repeats the previous behavior of its opponent. In contrast, indirect reciprocity theory implicitly assumes that the same individuals will not meet again (Nowak and Sigmund, 1998; 2005). Interactions are not repeated, but one shot. In particular, reputation is based on an image score that increases if a player cooperated in past interactions with other players and decreases if they did not do so. While in indirect reciprocity models cooperative behavior is based on how an individual has behaved against others, in direct reciprocity models cooperation rests on how an individual has behaved with the same partner in the past.

Suppose that advocates of the ELH manage to solve this issue and direct reciprocity provides the explanation for the evolution of human cooperation. Now according to Burnham's method, it is possible to predict our preferences from a conditional cooperative behavior like Tit-for-Tat[7]. However, this behavior does not single out a particular type of preference. For example, an evolutionary explanation based on Tit-for-Tat does not allow economists to suggest that cooperation is motivated by selfish preferences. While it can be the case that the preferences evolved were selfish, conditional cooperation in repeated interactions could also be served, among others, by a genuine altruistic preference (Trivers, 1971: p.51), an impure altruistic preference (Andreoni, 1990) or by a preference for fair outcomes. Hence, knowing that cooperation in our species most likely evolved by direct reciprocity does not provide a particular insight into what preferences have evolved in the ancestral environment.

This example points to the following problem. Although we might have reasons independent from direct reciprocity models to think that the underlying preferences are, for example, selfish, knowing that the correct explanation for the evolution of cooperative behavior is provided by direct reciprocity theory and that a conditional behavior like Tit-for-Tat evolved in the ancestral past provides no particular insight into human preferences. This might allow someone to claim that even if the modified version of Burnham's method generates a novel hypothesis that ends up being true after experimental research, it is only accidentally true. That is, it is true independently of the evolutionary model applied and this undermines the idea that Burnham's method can systematically provide new knowledge about human psychology.

The conclusion of this section is that hypotheses about human preferences are often inferred from an evolutionary model that focuses on fitness outcomes of behaviors. However, it is not possible to ascertain the presence of particular motives, preferences, desires or intentions based on such models. Put differently, it will routinely be possible to identify for some behavior and its ARE, multiple scientifically plausible hypotheses about human psychology. Based on such evolutionary models alone, it is hard to see how scientific progress could be made in

---

[7] To be clear, Tit-for-Tat is not always evolutionary successful in an indefinitely iterated Prisoner's Dilemma (e.g. Binmore, 1998)

determining which of these hypotheses are worth of experimental research. In the section that follows, I consider a way to refine Burnham's method and reduce the search space to a smaller number of hypotheses.

3. Evolutionary Models in Economics: Evolutionary Game Theory and the Indirect Evolutionary Approach

At the core of Burnham's method is evolutionary theory and this would be the focus in this section. Recall from the previous discussion that evolutionary researchers apply evolutionary models to determine the selection pressures for particular traits. Before evolution models are applied, one has to determine which traits are likely candidates for natural selection explanations. An open question remains: What kind of models should economists apply to infer hypotheses about preferences? We saw that knowing the right evolutionary model for some behavior does not tell economists what kind of preferences are more plausible from an evolutionary perspective. In what follows my aim is to present an alternative way of evolutionary theorizing that can successfully rule out all but a few hypotheses about human preferences.

It is possible to distinguish between two relevantly different types of evolutionary models. In one type the primary target of analysis is a psychological or cognitive trait internal to the organism. These models investigate evolutionary processes that produced the human psychologies that, in turn, contribute to human behaviors. Other types of models remain agnostic about human psychology and cognition and attempt to explain the evolution of particular behaviors.

To illuminate this distinction, in what follows I examine evolutionary game theory (henceforth EGT) and models of the IEA. What distinguishes the IEA from EGT is the shift of focus from behaviors to underlying preference patterns. EGT provides evolutionary explanations of behavior (or behavioral strategies). In contrast, models of the IEA explain the evolution of preferences directly and behavior indirectly. And as we will see, such models can narrow down the range of plausible hypotheses about human psychology that can be examined in the laboratory or the field.

3.1 Evolutionary Game Theory

Evolutionary research applies several approaches to modeling. Though one particular cluster of models can, however, be singled out: That of EGT. As we have seen in the previous section, Burnham's proposal is just an effort to hypothesize and predict preferences from evolutionary games. Models of direct or indirect

reciprocity theory are assumed to be correct, at least in their general outlines. Based on such models, economists attempt to predict which human preferences have evolved in past environments.

EGT and evolutionary dynamics were originally developed to explain why certain types of (supposedly genetically determined) behavioral strategies[8] are promoted by natural selection and stabilize in a population (e.g. Maynard Smith 1982). An evolutionary game refers to evolutionary processes such as natural selection and environment conditions such as repeated interactions or population structure that made a significant positive difference to the fitness of some organism that has a particular behavior[9] .

From an economist's perspective, what are the specific characteristics of evolutionary games? In a short but illuminating presentation of evolutionary models in sociobiology and game theory, D'Arms et al. (1998) observe that evolutionary game theorists seek to explain "behavior by pointing to adaptive advantages for those who engage in it, without attempting to explain how exactly tendencies to behave in the relevant way are embodied in a psychology" (D'Arms et al., 1998: p. 87). The question they attempt to answer is why a given type of behavior has evolved due to the winnowing process of selection, compared to many other behaviors that could have evolved.

Evolutionary game theorists, however, seek to explain types of behavior by examining the fitness benefits accruing to individuals, without providing an explicit account of preferences and decision-making that influence behaviors[10]. To be clear, it is recognized that it is important to understand the fine detail of human psychology (Trivers, 1971; Axelrod, 1984). However, cognitive or psychological traits are not an integral part of EGT. They are studied after building an evolutionary

---

[8] A strategy refers to "a behavioral phenotype; i.e., it is a specification of what an individual will do in any situation in which it may find itself" (Maynard Smith, 1982: p.10). In this chapter a strategy is "a way of behaving" (Kitcher, 1985: p.89), manifest behavior traceable to genes, often conditional to the "situation" an organism is in.

[9] Of particular importance to EGT are models that explore dynamic processes of Darwinian selection such as the replicator dynamic. However, dynamic processes often examined in EGT "do not build in any assumptions on behavior or knowledge, other than the basic principle of differential selection – apparently successful behavior increases its representation in the population, while unsuccessful behavior does not" (Mailath 1998: p. 1355).

[10] For example, according to Axelrod "there is no need to assume that the players are rational. They need not be trying to maximize their rewards. Their strategies may simply reflect standard operating procedures, rules of thumb, instincts, habits, or imitation" (Axelrod, 1984: p.18).

model and they are typically conceived as mere add-ons in the evolutionary explanation these models provide.

That said, this particular feature of EGT is also characteristic of much of evolutionary theorizing in economics. In a recent in-depth study on the evolution of human cooperation, Samuel Bowles and Herbert Gintis build evolutionary models based on parochialism, gene-culture coevolution theory and strong reciprocity (Bowles and Gintis, 2013). These evolutionary models focus on fitness and behavior, without examining questions of motivation and decision-making (ibid.: p. 168). In particular, they do not assume that individuals in a population face a choice between alternative options. Instead, evolution simply programs or "hard-wires" individuals with behavior conditional to norms. Moreover, these evolutionary models do not speak out on the issue of what passions, sentiments or neural mechanisms underlie social behaviors. In the ending chapters of their book, Bowles and Gintis present a distinct set of models that examines processes of norm internalization and social emotions like guilt and shame that bring about a particular type of behavior like altruistic punishment (Bowles and Gintis, 2013: Chapters10 and 11).

Let us take stock. In Burnham's method, models based on EGT are applied with the aim to provide novel hypotheses to explore in economic experiments. These evolutionary models focus on behavior, typically assumed to be genetically transmitted across generations. Moreover, the description of behavior in terms of human decision-making and human psychology is either very coarse or completely absent. To be clear, this does not make EGT trivial or empty of explanatory content. However, EGT is disconnected from any scientific programme that studies human decision-making and psychology. In the evolutionary explanations of EGT, nothing needs to be said (and in most cases, nothing is actually said) of how our mind functions. To determine what preferences follow from such evolutionary models, additional assumptions have to be made that are not part of the explanations these models provide.

It becomes also clear what kind of evolutionary theory can be more reliable in formulating hypotheses about human psychology. Evolutionary models that start from a more detailed description of human behavior in terms of decision-making and human psychology can also deliver a more refined description of the ARE. In turn, this can significantly reduce the range of plausible hypotheses about human psychology that can be systematically examined in the laboratory or the field.

3.2 The Indirect Evolutionary Approach

One way to overcome the limitations of EGT and refine Burnham's method is the following: economists can actively engage in evolutionary modeling and give evolutionary explanations of their own in which human psychology and decision-making receive their due. In fact, economists often analyze preferences and

decision-making from an evolutionary perspective, asking why we have particular psychological traits and seeking answers in terms of the contributions these traits have made to reproductive success. A case in point is the so-called Indirect Evolutionary Approach (IEA) initiated by Güth and Yaari (1992) and subsequently applied to investigate the evolution of other-regarding preferences (e.g. Akçay et al., 2009; Alger & Weibull, 2013, 2016; Heller and Mohlin, 2019).

The IEA aims to provide an evolutionary explanation of preferences patterns by asking what sort of selection pressures lead to their evolution and maintenance in a population. A robust mathematical result of the IEA is that self-regarding preferences, a standard assumption in much economic research, is not satisfactory from an evolutionary perspective. While social preferences often result in cooperative behavior and can be easily exploited, in the selection pressures formalized by these models, a combination of self-regarding and social preferences is more evolutionary stable than self-regarding preferences.

Let me give an example. In a series of papers Alger and Weibull build evolutionary models inspired by the work of Güth and Yaari (e.g. Alger and Weibull, 2013; 2016). They also assume that individual behavior is based on utility maximization, while evolutionary success is driven by payoffs related to reproductive success. They demonstrate that in simplified environments described by their models, natural selection leads to the evolution and maintenance of a certain family of preferences, convex combinations of selfish and social preferences. Their result is about the evolution and stability of preference types rather than behavior. A particular type of preference with a Kantian flavor called "Homo Moralis" has an evolutionary advantage. "Homo moralis" gives some weight to the principle of "acting according to that maxim whereby you can, at the same time, will that others should do likewise with some probability." (Alger and Weibull, 2016: p.57)

On a more technical note, in the IEA a large heterogeneous population consists of players with different preference types, modeled as parameters in utility functions. In accord with standard economic theory, players make decisions and calculate best responses, i.e. they are rational in that they maximize a utility function whenever called upon to make a choice. These players are matched in pairs and on the basis of the payoff structure of the game, preference types are inherited to the next generation. In the next period their offspring are matched to play the same game again. An evolutionary process based on fitness determines the long-run distribution of preference types in the population.

Some clarifications are necessary. In economics, the interpretation of utility and preferences is matter of dispute. Preferences and utility are typically given either a psychological (mental) or behavioral interpretation (e.g. Okasha, 2016; Dietrich and List, 2016). For some economists, preferences refer to psychological states inside the brain of decision makers. Others take preferences to be patterns of observable behavior. What must be emphasized for the IEA is that preferences and behavior are distinct theoretical constructs. In these models, a utility function is assigned to an individual player and a player's flexible choice of behavioral strategies depends on

preferences that mutate and transfer from one generation to another. In other words, in the IEA preferences refer to psychological traits manifested in behavior and these traits are acquired through genetic inheritance.

Moreover, fitness maximization is not an assumption of the IEA. If we take an IEA perspective and examine whether natural selection will favor or oppose psychological traits that guide human behavior, there is no need for a principle like (inclusive) fitness maximization. Models of the IEA distinguish between utility maximization driving the choice of behaviors and the fitness consequences of behavior. Based on economic theory, the assumption is that players will choose strategies to maximize their utility. However, utility differs from evolutionary fitness that determines the long-run survival of preferences in a population. While individuals may act rationally by choosing behavioral strategies that maximize utility, the behaviors chosen do not need to maximize reproductive success.

While there is much more that can be said about the IEA[11] , for present purposes it is enough to emphasize that IEA is different from EGT in the following way. In contrast to models that examine pre-programmed behavior, the IEA ascribes utility functions to players and models the evolution of preference types. The focus on preferences as the primary target of analysis and the assumption that decisions are made based on rational utility maximization is what distinguishes the IEA from EGT. However, EGT is not different from the IEA in other respects. What they share in common is theorizing based on tractable mathematical models, solution concepts like evolutionary stable strategies and stable fixed points as well as equations that describe evolutionary dynamics (like the replicator equation).

How can this type of evolutionary models refine Burnham's method and contribute to our understanding of human psychology? One way is by providing an evolutionary explanation. Applying models of the IEA can help economists explain psychological traits that are already known to be present in human populations. If we already have a good understanding of our psychology through laboratory and field experiments, evolutionary theorizing can help us determine the selection pressures that led to the evolution of psychological traits. Moreover, these models can contribute to economics even if we do not have yet a good grasp of human psychology. Determining the selection pressures might suggest hypotheses about additional features or details of the preferences included in the model. In both cases,

---

[11] An important feature of these models is that preference types are assumed to be (at least partially) observable. The degree to which these preferences are observed influences the outcome of preference evolution (e.g. Ok and Vega-Redondo, 2001). In their review of evolutionary models in economics, Robson and Samuelson (2011) consider this as a shortcoming of the IEA. For a recent model in which preferences and the cognitive capacity to conceal and reveal them co-evolves, see Heller and Mohlin (2019).

the IEA can help economists to find out more about traits that they are not completely ignorant about and are worth exploring further.

In particular, evolutionary models such as those of the IEA can contribute to the economic literature and the empirical study of preferences in the following ways.

First, when models based on the IEA suggest novel utility functions, like in the research of Alger and Weibull, the theoretical investigation into the evolutionary foundations of human sociality can be complemented by empirical research to find out whether participants in controlled laboratory experiments and the field actually have the psychological traits attributed to players by utility functions included in the model[12] .

Second, behavioral economics have developed mainly with the aim to build simple utility-based models to explain a wide variety of evidence in experimental research. These models expand the argument in utility functions to include a desire or preference for social outcomes. Inequity aversion and fairness utility functions are examples of such models (Fehr & Schmidt, 1999; Bolton &Ockenfels, 2000). In an experimental setting, it is possible to empirically examine how utility functions of evolutionary models fare against those typically suggested by other behavioral economists.

Third, by looking at a utility function it is often not immediately clear what kind of psychological processes underlie human decision-making. Preference in the IEA is an abstract concept that precedes choice and refers to a vast array of motivating factors (e.g. emotional commitments, aversions, desires). If, however, the explanations of these evolutionary models are found plausible, hypotheses can be formed about the details and properties that these preferences have. These hypotheses can be tested in the laboratory and the field. For example, neuroeconomics might provide a deeper understanding of human preferences by examining corresponding arrangements of neurons within the nervous system [13].

There is, however, an important issue that needs to be addressed. Economists that apply the IEA do not make an attempt to investigate whether the selection pressures described by their models reflect conditions present in past environments. In evolutionary theory the relevant context is historical, namely the ARE that impinged upon particular animal and human traits. The emphasis however of the IEA (and many models of EGT) is on the mathematical aspects of evolutionary theory. In

---

[12] An interesting application of the IEA can be found in the work of Clavien and Chapuisat (2016). They investigate the details of the argument contained in a family of utility functions advanced by Alger and Weibull and conclude that genuine altruistic motivations can be found in long term reciprocal interaction and.kin and romantic relationships.

[13] For a discussion of different projects in neuroeconomics, see Vromen (2011).

these models the result "springs out from the mathematics" (Alger and Weibull, 2013: p.2271) and "mathematics show the way from evolutionary stability conditions to the preferences that evolution favors" (Alger and Weibull, 2016: p.57). What is problematic is that evolutionary scientists do not draw upon historical information to build their evolutionary models. Moreover, they do not examine empirical data that could confirm or invalidate the assumptions postulated by their models. In short, these models are ahistorical in the sense that relevant historical facts are completely absent from the evolutionary explanations they provide.

The point here is that researchers not solely concerned with purely theoretical problems internal to evolutionary theory have to provide evidence that support the model construction. By disregarding historical data in studying the workings of selection processes, theorists run the risk of building evolutionary models that leave out important variables, include irrelevant variables or fail to account for the actual functional relationship between independent and dependent variables. It is not only important that their evolutionary models are mathematically coherent, but they must also be empirically adequate in that the selection pressures advanced must have actually operated over a stretch of human evolutionary past[14] . If the assumptions of evolutionary models are largely inaccurate, the explanations they provide will not represent the historical processes they are supposed to examine.

A serious attempt to check whether there are historical studies that support evolutionary models of cooperation can be found in the work of Bowles and Gintis (Bowles and Gintis, 2013: Chapter 6; see also Sterenly, 2014). They examine in detail historical evidence to make clear that their models are consistent with the actual process that they purportedly study. For example, they present evidence of intergroup conflict and high rates of violent death in the late Pleistocene, to advance the view that human cooperation evolved by group selection. Moreover, they examine the size of nomadic hunter-gather bands and conclude that in contrast to what is typically assumed, they were not small but included around 90 members. Based on such evidence, they claim that simple direct reciprocity models do not adequately explain the evolution of human cooperation. It is unfortunate that advocates of the IEA have not carried out a similar analysis of the conditions allegedly obtained during preference evolution.

To sum up, the IEA points out to a direction that evolutionary minded economists can take, should they be interested in applying evolutionary method. Most of the work in EGT and sociobiology ignores human psychology and decision-making, focusing on behaviors and their fitness effects. In contrast, economists often apply the IEA and assume that players are utility maximizers. Moreover, they treat human

---

[14] An enduring criticism of explanations that evolutionary researchers typically give for traits is that they are analogous to Kipling's "Just So Stories" (Gould &Lewontin 1979). This criticism refers to accepting evolutionary explanations without sufficient support by empirical evidence.

preferences as a type of biological trait that is acquired, modified and retained by outcompeting other traits in the population. Based on the discussion of the IEA, my argument is that instead of evolutionary models that privilege patterns of behavior, economists can draw upon or build evolutionary models where preferences and choice take center stage. Such evolutionary models can constrain the hypotheses space because they screen off those preference types that are not evolutionary successful.

4. Decision-Making and Utility Maximization from an Evolutionary Perspective

In the previous section I argued that Burnham's method will be more reliable if economists apply evolutionary models that include some assumptions about human psychology and decision-making. This refinement of Burnham's method was based on the IEA, a theory devoted to studying the evolution of preferences. Moreover, I argued that a model's specifications have to be borne out from or have to be supported by empirical evidence about past environments. Hence, the following problem emerges. By design, these models include a decision-making process, typically based on rational utility maximization. One might wonder whether this assumption is plausible, given what we know about our ancestral past. Moreover, there is no shortage of alternative approaches that can be found in the literature (e.g. dual process theory, heuristics and biases, social learning biases, massive modularity hypothesis). An economist might want to consider substituting the utility maximization component of the IEA with a different model of decision-making.

In the IEA, economists' main interest is not whether people are rational utility maximizers, as most economic models typically assume, but to investigate the empirical and theoretical relevance of self-regarding and social preferences, given that people are rational. These models take for granted that players act rationally and make choices that maximize a utility function. Although utility maximization has been successfully applied in economics, it is often criticized on theoretical, empirical and conceptual grounds. However, the economists discussed in the previous section apparently hold that the assumption of utility maximizing players can be retained when taking insights from evolutionary theory into account. What evolutionary theory necessitates is modifying the arguments included in utility functions (e.g. Robson and Samuelson, 2011; Bowles and Gintis, 2013).

Not every economist agrees with this view. Ben-Ner and Putterman (1998; 2000) for example, explore what evolutionary theory has to offer to economics. Tellingly, they write that "in the long-run, the economics discipline as a whole will recognize that the old assumption of rational, strictly self-interested individuals is not only an inexact and special approximation, but also inconsistent with a scientific view of human nature as the product of an evolutionary process" (2000; p.97). Roughly

speaking, more radical economists think that evolutionary theory implies that the standard economic theory should be discarded altogether (Burnham et al, 2016: p.139). And they often put forward alternative decision-making accounts such as the "adaptive toolbox" approach with fast and frugal heuristics (e.g. Gigerenzer and Selten, 2001) or modularity-based theories (e.g. Kurzban, 2010)[15].

What does evolutionary theory has to say about human decision-making? From an evolutionary perspective, is there a plausible account of decision-making? Unfortunately, there are no clear answers to these questions. It is worrying that very little work has been done to develop evolutionary arguments, build evolutionary models and test claims about human decision-making. Economists of the IEA, for example, have not provided arguments for the implicit claim that utility maximizers instead, for example, of those that apply heuristics and biases must have had an evolutionary advantage. What they examine is whether some features of utility functions (i.e. preference types) would outperform in terms of reproductive success competing features of utility functions. Hence, it remains an open question whether (and to what extent) such decision-making capacities evolved in response to challenges of past environments.

When economists present intuitive evolutionary arguments in favor of utility maximization, the result remains inconclusive. Take for example Robson's (2001) claim that the evolutionary function of expected utility is that it allowed optimal, rapid and flexible responses to novel environments[16] . To the extent that humans make utility maximizing choices, they do so because they are endowed with some sort of a general-purpose cognitive device. In constantly changing environments, such a device to make relatively accurate, fast and flexible decisions among alternative options might have been beneficial to survival and reproduction.

There are, however, alternative decision-making accounts informed by evolutionary theory. To give one example, a group of researchers around Gerd Gigerenzer claim that in our decision-making process we are using simple satisficing, domain-specific heuristic rules (Gigerenzer&Selten, 2001). Although this group does not engage in evolutionary modelling, it purports that heuristics have evolutionary origins. They write that "evolution would seize upon informative environmental dependencies [....] and exploit them with specific heuristics if they would give a decision-making organism an adaptive edge" (Gigerenzer& Todd, 1999: p. 19). What Gigerenzer and

---

[15] Some economists believe that the neuroscientific investigation of the human brain will eventually lead to new mechanistic theories of decision-making (e.g. Burnham et al 2016). While I do not doubt that such research can materialize in the future, in this section I only examine decision-making accounts currently applied by economists in their research.

[16] For an in-depth examination of Robson's evolutionary project, see Vromen (2003).

the ABC group appear to emphasize is that in a relatively stable environment with recurrent but distinct evolutionary problems, relying on simple heuristics would have been adaptive because of the benefits in decision-making speed and low energy expenditure.

This quick overview by no means exhausts all the accounts expounded about human decision-making in economics. It suffices, however, to make the following point: It is not always evolutionary advantageous to make decisions according to standard economic theory. In a novel environment, having a general-purpose device might have been adaptive since it would allow for flexible choices that match new conditions better than behavior led by fixed rules. A constantly changing environment would render fixed heritable rules obsolete after a few generations. That, though, is not enough to exclude heuristics from being adaptive in the ancestral environment. Under more stable conditions, decisions according to rigid but computationally simple heuristic rules might have outperformed in terms of survival and reproduction decisions based on utility maximization.

The issue here that economists cannot simply indicate what kind of environments might have favored a particular decision-making theory. And it is not possible to develop their line of argument any further without an insight about the actual selection pressures obtained in our ancestral past. To my knowledge, economists that build models based on the IEA have not made a serious attempt to provide empirical evidence that is relevant to assess whether our past speaks in favor of utility maximization, or against it. Without a detailed understanding of the environment in which alternative decision-making capacities evolved, it is not possible to argue as to whether decisions based on utility maximization outperformed simple heuristics. It is not even possible to make a particularly strong case in favor of the weaker claim that decision-making based on standard economic theory was significantly common in past environments. The upshot is that it remains unclear to the extent that humans behaved rationally because they were endowed with some sort of a general-purpose utility maximizing device or their behaviors availed a multitude of special-purpose heuristic rules.

How can one determine which decision-making account is the most likely candidate to be included in models that provide natural selection explanations? The best means of approaching this issue is to supplement evolutionary arguments with evidence from fields such as primatology that focuses on the characteristics of extant non-human primates and ancestral hominids and anthropology that examines specific hunter-gatherers groups (Marlowe, 2005). Moreover, phylogenetic analysis for hominid evolution could provide a chronology of major evolutionary changes in the social environment (Foley, 1996). Although evidence showing what kind of decision-making must have been operative in our ancestral past is by no means easy to come by, evolutionary minded economists could put some research effort into gathering data that would serve as the basis for exploring the fitness benefits and costs of alternative decision-making capacities.

An important clarification to this argument needs to be made. From an evolutionary perspective, preferences are not at issue here. The problem is that economists of the IEA take for granted that decisions are made by a cognitive device that leads to utility maximization. The same point will still apply if the same economists routinely assumed that human decision-making is based on domain-specific modules, heuristics and biases, social learning biases, a dual process system or any other human decision-making process. The difference would be that, in this alternative case, they would have ignored to consider in their evolutionary models competing decision-making theories, which nevertheless could have found support by empirical evidence from anthropology, archaeology or primatology.

To conclude this section, it turns out that the criticism raised against the utility maximization assumption of the IEA also applies to evolutionary arguments for or against any particular account of decision-making. Whether evolution favors domain specific modules, simple heuristics, dual process theory or utility maximization depends on the ARE. From an evolutionary perspective, it is not possible to assess alternative decision-making theories without an empirical analysis of specific features present in multiple past environments. In addition to abstract evolutionary theorizing, more empirical work is necessary to unravel the ARE against which many of our traits evolved.

5. Conclusion

Let us recap. For Burnham it is possible to formulate hypotheses and make predictions about human psychological traits, on the basis of what patterns of behavior would have likely been adaptive in the ARE. I argued that Burnham's method is evolutionary in the sense that evolutionary models are applied in determining the ARE for a particular behavior, and subsequently in searching for and framing hypotheses about psychological traits that can be explored in economic experiments. As we have seen, there will routinely be several scientifically plausible psychological hypotheses compatible with any given evolutionary model of behavior, and this raises the problem of grounding specific hypotheses about human psychology into such models.

This however should not lead economists to question the value of evolutionary theory when applied to the social sciences. Economists' interest in the details of psychological traits and decision-making has brought their evolutionary theorizing far-off to that of EGT and sociobiology. Burnham's method would be more reliable if instead of applying evolutionary models that target behavior, economists build evolutionary models of their own that target particular features of human psychology and cognition. In particular, economists of (or close to) the mainstream might opt for evolutionary models of the IEA that allow for an enrichment of received economic theory with novel insights about human preferences. Such

models would allow economists to make progress in elucidating traits related to human psychology.

The drawback is that economists have to provide sufficient reasons why their evolutionary models commit to a particular account of decision-making. We saw that it is still an open question which decision-making theory is plausible from an evolutionary perspective. The proliferation of evolutionary models in economics did not result in significant contributions to the collection and processing of data that can illuminate the environment in which many of our traits evolved. It is important that an evolutionary study of human preferences and decision-making should also make use of empirical facts about the actual selection pressures at work during human evolutionary history.

References

Akçay, E., Van Cleve, J., Feldman, M. W., &Roughgarden, J. (2009). A theory for the evolution of other-regard integrating proximate and ultimate perspectives. Proceedings of the National Academy of Sciences, 106(45), 19061–19066. https://doi.org/10.1073/pnas.0904357106

Alger, I., and Weibull, J. W. (2013). Homo Moralis: preference evolution under incomplete information and assortative matching. Econometrica, 81: 2269-2302.

Alger, I., & Weibull, J. W. (2016). Evolution and Kantian morality. Games and Economic Behavior, 98, 56–67. https://doi.org/10.1016/j.geb.2016.05.006

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. The Economic Journal, 100(401), 464. https://doi.org/10.2307/2234133

Axelrod, R. (1984). The Evolution of Cooperation. New York: Basic Books.

Ben-Ner, A., &Putterman, L. (1998). Values and institutions in economic analysis. In: Ben-Ner, A., Putterman, L. (Eds.), Economics, Values and Organization. Cambridge University Press, New York, p. 3–69.

Ben-Ner, A., &Putterman, L. (2000). On some implications of evolutionary psychology for the study of preferences and institutions. Journal of Economic Behavior & Organization, 43(1), 91–99. https://doi.org/10.1016/s0167-2681(00)00110-4

Binmore, K. (1998). Review of R. Axelrod's "The complexity of cooperation: Agent-based models of competition and collaboration". Journal of Artificial Societies and Social Simulation, 1(1).

Binmore, K., & Shaked, A. (2010). Experimental economics: Where next? Journal of Economic Behavior & Organization, 73(1), 87–100. https://doi.org/10.1016/j.jebo.2008.10.019

Bolton, G. E., &Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. American Economic Review, 90(1), 166–193. https://doi.org/10.1257/aer.90.1.166

Bowlby J. (1969). Attachment and Loss. Volume I: Attachment. Basic Books, New York.

Bowlby, J. (1973). Attachment and Loss. Volume II: Separation, Anxiety and Anger. Basic Books, New York.

Bowles, S., & Gintis, H. (2013). A cooperative species : human reciprocity and its evolution. Princeton University Press.

Burnham, T. C. (2013). Toward a neo-Darwinian synthesis of neoclassical and behavioral economics. Journal of Economic Behavior & Organization, 90, S113–S127. https://doi.org/10.1016/j.jebo.2012.12.015

Burnham, T. C., & Hare, B. (2007). Engineering Human Cooperation. Human Nature, 18(2), 88–108. https://doi.org/10.1007/s12110-007-9012-2

Burnham, T.C., & Johnson, D. (2005). The biological and evolutionary logic of human cooperation. Analyse and Kritik 27, 113–135

Burnham T.C., Stephen E. G., Lea S. E., Bell A., Gintis H., Glimcher P.W., Kurzban R., Lades L., McCabe K., Panchanathan K, Teschl M., and Witt U. (2016). Evolutionary Behavioral Economics, Wilson D.S. and A. Kirman, eds. Complexity and Evolution: Toward a New Synthesis for Economics. Strungmann Forum Report 19, Cambridge, MA: The MIT Press.

Clavien, C., &Chapuisat, M. (2016). The evolution of utility functions and psychological altruism. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 56, 24–31. https://doi.org/10.1016/j.shpsc.2015.10.008

Collins, J., Baer, B., & Weber, E. J. (2016). Evolutionary Biology in Economics: A Review. Economic Record, 92(297), 291–312. https://doi.org/10.1111/1475-4932.12260

D'Arms, J., Batterman, R., &Gorny, K. (1998). Game Theoretic Explanations and the Evolution of Justice. Philosophy of Science, 65(1), 76–102. https://doi.org/10.1086/392627

Dietrich, F., & List, C. (2016). Mentalism versus Behaviourism in Economics: A Philosophy of Science Perspective. Economics and Philosophy, 32(2), 249–281. https://doi.org/10.1017/s0266267115000462

Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. The Quarterly Journal of Economics, 114(3), 817–868. https://doi.org/10.1162/003355399556151

Fehr, E., & Schneider, F. (2009). Eyes are on us, but nobody cares: are eye cues relevant for strong reciprocity? Proceedings of the Royal Society B: Biological Sciences, 277(1686), 1315–1323. https://doi.org/10.1098/rspb.2009.1900

Foley, R. A. (1996). An evolutionary and chronological framework for human social behavior. Proceedings of the British Academy 88: 95-117.

Gigerenzer, G & R Selten. (2001). Bounded rationality : The adaptive toolbox. The Mit Press.

Gigerenzer, G. & Todd, P. M. & the ABC Research Group (1999). Simple heuristics that make us smart. Oxford University Press, , Cop.

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 B, 205(1161), 581–598

Grafen, A. (2006). Optimization of inclusive fitness. Journal of Theoretical Biology, 238(3), 541–563. https://doi.org/10.1016/j.jtbi.2005.06.009

Grafen, A. (2014). The formal darwinism project in outline. Biology & Philosophy, 29(2), 155–174. https://doi.org/10.1007/s10539-013-9414-y

Griffiths, P. E. (1996). The Historical Turn in the Study of Adaptation. The British Journal for the Philosophy of Science, 47(4), 511–532. https://doi.org/10.1093/bjps/47.4.511

Güth, W., &Yaari, M. (1992). An evolutionary approach to explain reciprocal behavior in a simple strategic game. In U. Witt (Ed.), Explaining Process and Change - Approaches to Evolutionary Economics. Ann Arbor: University of Michigan Press.

Güth, W. (1995). An evolutionary approach to explaining cooperative behavior by reciprocal incentives. International Journal of Game Theory, 24(4), 323–344. https://doi.org/10.1007/bf01243036

Hagen, E. H., & Hammerstein, P. (2006). Game theory and human evolution: A critique of some recent interpretations of experimental games. Theoretical Population Biology, 69(3), 339–348. https://doi.org/10.1016/j.tpb.2005.09.005

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. Evolution and Human Behavior, 26, 245–256.

Heller, Y., &Mohlin, E. (2019). Coevolution of deception and preferences: Darwin and Nash meet Machiavelli. Games and Economic Behavior, 113, 223–247. https://doi.org/10.1016/j.geb.2018.09.011

Hirshleifer, J. (1977). Economics from a Biological Viewpoint. The Journal of Law and Economics, 20(1), 1–52. https://doi.org/10.1086/466891

Huck, S. (1997). Institutions and Preferences: An Evolutionary Perspective, Journal of Institutional and Theoretical Economics, 153(4): 771-779

Irons, W. (1998). Adaptively relevant environments versus the environment of evolutionary adaptedness. Evolutionary Anthropology: Issues, News, and Reviews, 6(6), 194–204. https://doi.org/10.1002/(sici)1520-6505(1998)6:6<194::aid-evan2>3.0.co;2-b

Kanazawa, S. (2003). The Savanna Principle. Managerial and Decision Economics, 25(1), 41–54. https://doi.org/10.1002/mde.1130

Kitcher, P. (1985). Vaulting ambition : sociobiology and the quest for human nature. MIT Press.

Kurzban, R. (2012). Why everyone (else) is a hypocrite : evolution and the modular mind. Princeton University Press.

Lehmann, L., Alger, I., & Weibull, J. (2015). Does evolution lead to maximizing behavior? Evolution, 69(7), 1858–1873. https://doi.org/10.1111/evo.12701

Mailath, G. J. (1998). Do people play Nash equilibrium? Lessons from evolutionary game theory. Journal of Economic Literature 36(3), 1347–1374.

Marlowe, F. W. (2005). Hunter-gatherers and human evolution. Evolutionary Anthropology: Issues, News, and Reviews, 14(2), 54–67. https://doi.org/10.1002/evan.20046

Maynard, J. (1982). Evolution and the theory of games. Cambridge Cambridge Univ. Press.

Nettle, D., Harper, Z., Kidson, A., Stone, R., Penton-Voak, I. S., & Bateson, M. (2013). The watching eyes effect in the Dictator Game: it's not how much you give, it's being seen to give something. Evolution and Human Behavior, 34(1), 35–40. https://doi.org/10.1016/j.evolhumbehav.2012.08.004

Ok, E. A., & Vega-Redondo, F. (2001). On the Evolution of Individualistic Preferences: An Incomplete Information Scenario. Journal of Economic Theory, 97(2), 231–254. https://doi.org/10.1006/jeth.2000.2668

Okasha, S. (2016). On the Interpretation of Decision Theory. Economics and Philosophy, 32(3), 409–433. https://doi.org/10.1017/s0266267115000346

Richerson, P. J., & Boyd, R. (2006). Not by genes alone : how culture transformed human evolution. The University Of Chicago Press.

Robson, A. J. (2001). Why Would Nature Give Individuals Utility Functions? Journal of Political Economy, 109(4), 900–914. https://doi.org/10.1086/322083

Robson, A.J., Samuelson, L. (2011). The evolutionary foundations of preferences. In: Benhabib, J., Bisin, A., Jackson, M. (Eds.), The Social Economics Handbook. North Holland, Amsterdam, pp.221–231

Sterelny, K. (2014). Cooperation, culture, and conflict. The British Journal for the Philosophy of Science, 67(1), 1–31.

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. The Quarterly Review of Biology, 46(1), 35–57. https://doi.org/10.1086/406755

Vromen, J. (2003). Why the economic conception of human behaviour might lack a biological basis. Theoria, 18(48), 297-323.

Vromen, J. (2011). Neuroeconomics: Two camps gradually converging. What can economics gain from it? International Review of Economics, 58. 267-285.

Chapter 3:Sexual Selection of Conspicuous Consumption[17]

1. Introduction

Conspicuous consumption is one of the most striking phenomena in recent human history. Thorstein Veblen (1899) used the concept to bring forward original and provocative insights into the relation between consumer behavior and social status. Whereas Veblen mainly focused on the elite of the middle to late 19th century western societies, conspicuous consumption is not limited to particular social groups but has an enormous diffusion to a vast part of contemporary human population. Over the past decades, consumers belonging in different economic, religious, ethnic and linguistic groups are increasingly engaged in conspicuous consumption.

Why has conspicuous consumption evolved? What are the psychological mechanisms that underlie conspicuous consumption? And can we relate human reproductive decisions with the evolution of a culture that values and rewards conspicuous consumption?

Answering these questions is a daunting task. Sociological, biological, psychological and economic factors are deeply intertwined in the process of conspicuous consumption and it is difficult to completely isolate them from each other. Moreover, conspicuous consumption is a phenomenon driven by social interactions embedded in cultural groups, socioeconomic classes and communities in different regions and countries. Hence, it is difficult to assert the primacy of particular causal factors.

In this chapter I focus on the evolutionary explanation of conspicuous consumption, since much has been written on this topic in the last two decades. An increasing body of research attempts to explain consumer behaviors by applying evolutionary insights and methods (Miller 2001 and 2009; Saad 2014, Witt 2016). The current trend in the evolutionary literature is the emphasis on explaining conspicuous consumption by applying sexual selection. A number of papers draw upon ideas from costly signaling theory and sexual selection to argue that conspicuous consumption has evolved as a sexually selected mating strategy that provides reproductive benefits for males (e.g. Zak and Park 2002, Griskevicius et al. 2009, De Fraja 2009, Nelissen and Meijers 2011, Sundie et al. 2011, Collins Baer and Weber 2015). Like peacocks' large and colorful tails, conspicuous consumption impedes males' chances of survival while it confers mating benefits.

---

[17] A paper based on this chapter is currently under peer review.

In what follows, I examine the arguments put forward by these evolutionary researchers and I identify areas of controversy that may spark a debate among economists, psychologists and evolutionary biologists. Though I share their insight that evolutionary theory can contribute to the explanation of consumption practices, I argue that in the more recent literature evolutionary researchers attempt to examine the evolutionary underpinnings of conspicuous consumption theory without showing any interests in analyzing evidence about the ancestral environment, psychological features or socio-cultural parameters. In particular, the evolutionary explanation of conspicuous consumption based on sexual selection faces three problems:

- First, there is the problem of flexibly defining conspicuous consumption: does this trait refer to psychological capacities that we have or certain types of behavior in which we engage?
- Second, there is the problem of whether conspicuous consumption has evolved as a sexually selected mating strategy: Conspicuous consumption does not conform to theconditions for the reliability of costly signalling traits.
- Third, there is the problem of the type of evidence needed to support an evolutionary explanation based on sexual selection: Available ethnographic evidence from foraging societies cast doubt whether conspicuous consumption evolved as a mating strategy.

Given these problems, in this chapter I argue for an alternative process that could explain the evolution of conspicuous consumption. In particular, cultural evolution theory provides a useful way to study consumption behaviors in modern environments. A plausible answer to the question of why people engage in conspicuous consumption involves cultural transmission: people observe and learn the behaviors of others during social interactions. The implication is that conspicuous consumption does not evolve as a display to potential mates but as an effect of pre-existing evolved psychological adaptations.

Cultural evolution theory has not made inroads into consumption research, despite its growing representation in modern evolutionary research. This is unfortunate, given the interest of evolutionary researchers in examining cultural processes and their aspirations to explain human cultural traits. In this chapter I apply cultural evolution theory to explain conspicuous consumption and I critically compare it with alternative approaches taken by evolutionary researchers. Given the novelty of applying cultural evolution to the study of consumer behavior and to facilitate future research, I focus on the motivating role of social learning biases in directing consumers towards their decisions.

The chapter is structured as follows. In the next section I provide a brief and general introduction to sexual selection, parental investment and costly signaling theory, focusing on the aspects I consider more relevant for the study of contemporary conspicuous consumption. I outline what are considered to be the criteria for

arguing that a trait is the outcome of sexual selection. In section 3, I examine what kind of traits evolutionary researchers have in mind by conspicuous consumption. Section 4 explores in more depth whether conspicuous consumption does in fact possess the main features of sexual adaptations. Following this, in section 5 I review and evaluate the type of evidence needed to support sexual selection explanations. This section deals with evolutionary history and the importance of evidence about the ancestral environment to support hypotheses about sexual adaptations. Section 6 explores areas of research where cultural evolution can provide an original contribution with its commitment to develop hypotheses informed by psychology and evolutionary theory. I focus on cultural transmission biases that influence the propagation of conspicuous consumption in modern environments. Section 7 concludes.

2. Evolutionary Foundations: Sexual Selection, Parental Investment and Costly Signalling Theory

A number of recent papers draw upon ideas from sexual selection, parental investment and costly signalling theory to argue that conspicuous consumption is a sexual adaptation (Griskevicius et al. 2007, Miller 2009, Fraja 2009, Nelissen and Meijers 2011, Sundie et al. 2011, Collins, Baer and Weber 2015). In what follows I briefly describe the insights that these theories provide and then explain how the process of sexual selection can lead to the evolution of traits.

Sexual selection is often invoked to explain the evolution of exaggerated costly traits. It is an evolutionary process that results in differential mating success and can generate rapid genetic and phenotypic change (Darwin 1871, Maynard-Smith 1978). Sexual selection comes in two main forms: a) intrasexual selection by competition (usually among males) and b) intersexual selection by (usually female) choice. Intrasexual selection leads to exaggerated traits that promote the reproductive success in species where members of the same sex compete for access to mates of the other sex. A classic example is the large antlers of male red deer that repel rivals and allow access to territory and female herds. In intersexual selection,members of one sex choose as a mate a member of the opposite sex. This leads to the evolution of traits that make an organism more attractive as a mating partner for members of the opposite sex.

The fact that especially males tend to engage in signaling to attract sexual partners is often assumed to result from the asymmetry in parental investment among males and females - the contribution in resources (i.e. time, energy) of each parent in the production of viable offspring (Trivers 1972). Due to gestation and nursing, females often provide parental investment that is significantly higher than males. Female high parental investment makes them reproductively valuable for males and female choice is more pronounced than male choice. Males attempt to convince females of

their qualities as mating partners and they often compete with each other for women's sexual interest.

So far I have described the way sexual selection and parental investment is typically presented in literature. I now will briefly describe the main problem with explanations based on sexual selection and the way costly signalling theory provides a solution.

A key characteristic of sexual selection is that it results in traits to evolve even when they do not benefit or even inhibit an organism's survival. Put in a different way, natural selection may act against sexual selection when females choose traits that have a negative impact on male survival and pass these traits to their male offspring. Those females that prefer males that do not have traits detrimental to the male's survival will produce offspring that live longer and will have more sexual interactions. Hence the challenge for evolutionary research is to provide an evolutionary explanation of why extreme traits do better than more moderate traits.

A substantial literature on signaling theory in animal behavior proposes the following solution to this problem (Zahavi 1975, Grafen 1990a, b, Bliege Bird and Smith 2005). Some extravagant traits are good indicators of particular male qualities that females find important in their choice of a mating partner. The relation between the expression of a relatively costly male trait and particular male qualities is such that only higher-quality males can afford to express it. Hence, a female can use this trait as an indicator of a male's genetic or phenotypic qualities. Females benefit if they choose males who display an extreme trait, while males compensate the cost of the display by an increase in mating opportunities.

In the above evolutionary explanation it is possible to distinguish the following traits:

i) A male display trait, a costly quality-dependent signal emitted by human males;
ii) Male quality trait(s) indicated by the display trait;
iii) A female's preference[18] for males with better quality traits. Since male quality traits are often unobservable, females prefer males with higher levels or better quality of the display trait;

Next I will examine how evolutionary researchers examine and propose to support this claim. Costly signalling theory has a paradigm-like status in evolutionary research, and continues to serve as the main resource for many different types of

[18] To keep the analysis simple, I assume that a female's preference for a mating partner coincides with her mating choice. This is not the case when there is competition among females to attract mates and males demonstrate a preference for a female with certain characteristics (e.g. resources, attractiveness).

work in evolutionary psychology, evolutionary behavioral economics and human behavioral ecology. Griskevicius et al (2007) and Nelissen and Meijers (2011) present the conditions for a trait to qualify as a costly display:

i) it must be easily observable by others;
ii) it must be hard to fake because of its associated costs to the signaler in terms of economic resources, time, energy, risk, whereby the costlier the trait the more likely it is to be an honest indicator;
iii) it must result in a fitness benefit for the male signaller such as an increased access to desirable female mates;
iv) it must be an indicator to potential female mates of some unobservable but desirable male trait or characteristic, such as access to resources, prosocial disposition, courage, physical health or intelligence;

The evolutionary argument is that human conspicuous consumption has just these four characteristics and therefore it is a sexual adaptation. Hence, there are two ways to criticize an evolutionary explanation of conspicuous consumption based on sexual selection and costly signalling theory. First, it is possible to question the claim that conspicuous consumption does have these four features. Second, we can demonstrate that there is an alternative evolutionary explanation that does not result in problems revealed by a detailed examination of these four key characteristics. In the next sections I will follow these two lines of argumentation. Before presenting my arguments however, it is important to clarify what kind of traits these researchers have in mind when they use the term conspicuous consumption.

3. What is the main trait in question?

One of the main problems in this literature is that it is often hard to understand what kind of trait do evolutionary researchers refer to by conspicuous consumption.A close reading of evolutionary models and empirical research suggests that a single term covers different concepts and processes across many disciplines. In the section that follows I will examine in more detail evolutionary explanations based on sexual selection in an attempt to clarify the main trait under investigation.

In the presentation of a model inspired by costly signalling theory, Fraja (2009) states that it is the psychological desire for conspicuous consumption that is designed by sexual selection (p.62). For Collins, Baer, & Weber (2015) conspicuous consumption refers to a male preference that has deep evolutionary roots. Yet, in

other parts of his paper, Fraja (2009) defines conspicuous consumption as *"consumption for its own sake"* that "*is easy to observe and expensive to acquire"* (p.52). In Collins, Baer, & Weber (2015) conspicuous consumption includes *"any consumption activities beyond those required for survival"* such as *"developing art or other objects of beauty in traditional societies, or participating in the labour force to earn income in modern times"* (p.191). Here it is assumed that a minimum level of consumption is necessary for survival (e.g. health care, food, housing) and all male activities that do not contribute to survival (e.g. entertainment, travelling, painting) are conspicuous consumption. Griskevicius et al. (2007) and Sundie et al (2011) identify themselves as evolutionary psychologists and argue that in their experimental study they investigate the design features of psychological adaptations for conspicuous consumption (e.g. Griskevicius et al., 2007: p.99). Their research focuses on some sort of motivational aspects predicted by costly signalling theory.

What can we conclude from this multitude of different formulations of conspicuous consumption? Conspicuous consumption is often treated as a psychological trait that is the target of sexual selection. In evolutionary research stemming from economics, conspicuous consumption refers to individuals' subjective motivations (desires, intentions, beliefs or preferences) that underlie behavior. A closely related understanding of conspicuous consumption in evolutionary psychology refers to actual causal mechanisms that produce behavior. What sexual selection explains is not human behavior but their behavior generating mechanisms that we have. Put differently, sexual selection produces behavior only indirectly, through psychological mechanisms.

The key problem with these alternative formulations of conspicuous consumption is that they are insufficiently clear to allow an in-depth understanding of the trait under investigation. Evolutionary researchers frequently use the term psychological mechanism to refer to everything from input-output information processes to subjective mental states involved in conspicuous consumption. Although it is common for evolutionary psychologists and economists to examine the design characteristics of our brain system, in this literature they are not explicit on the model of human cognition and psychology they apply. Does our brain contain specialized mechanisms or "modules" that evolved to solve fitness-relevant problems in the ancestral environment? Or is our brain a general-purpose problem solving device? It is not clear whether many of these conceptions of conspicuous consumption are rivals or they complement each other.

A distinct formulation is often employed by evolutionary modellers to analyze conspicuous consumption. In evolutionary theory, conspicuous consumption is understood as behavior in relation to fitness outcomes, independently of the actor's conscious or subjective motivations. What sexual selection explains is why organisms acquired the behaviors that they have in the present environment. For example, in costly signalling theory behaviors and behavioral strategies are often described as adaptive traits. What these models examine is fitness consequences of types of behaviors that are present in the human population.

The upshot is that there are different ways to think about conspicuous consumption that diverge from one another at the conceptual level. To understand the phenomenon of conspicuous consumption, I think it is essential to disentangle behavioral outcomes from motivational aspects. Conspicuous consumption can be a kind of a) internal behavior-generating mechanism or b) overt behavior produced by a mechanism[19]. Hence, it is important for researchers to be clear about what exactly is the main display trait they examine in their studies. Do they suggest that conspicuous consumption is a type of behavior that requires an evolutionary explanation? Or is conspicuous consumption identified as psychological trait that interacts with the environment to produce particular behavior?

In what follows I consider conspicuous consumption as a type of behavior. It is uncontroversial from the perspective of costly signalling theory and sexual selection that conspicuous consumption is not marked by specific subjective states within individuals. Evolutionary theorists rarely describe their models as comprising psychological mechanisms. However, these evolutionary models of behavior and strategies are often applied to generate testable hypotheses about human psychology. Although costly signalling theory and sexual selection examine types of behavior, they are an indispensable starting point for more sophisticated explorations of psychological hypotheses that are subjected to empirical test.

In the next sections I follow this line of argumentation: first, I examine whether the evolutionary explanations of conspicuous consumption meet the conditions specified in the literature on animal signaling theory and sexual selection and second, I analyze the type of evidence needed to support an evolutionary explanation based on sexual selection.

4. Conspicuous Consumption and Costly Signalling Theory

Whatever traits different researchers have in mind, they have to actually possess the four defining features presented in Section 2. In order for evolutionary researchers to support their argument, they must offer evidence *that conspicuous consumption is an observable, costly trait that indicates particular qualities and provides fitness benefits to those that possess it*. In the following sub-sections I argue that there are serious problems with this claim.

---

[19]An alternative is presented in the experimental study of Nelissen and Meijers (2011). They examine the effects of luxury goods and argue that wearing specific brand-labelled clothing meet the criteria of Bird and Smith (2005) and qualify as costly signals. In this chapter I will not examine the idea that material goods can be human sexual adaptations.

4.1 Conspicuous Consumption is Observable

Conspicuous consumption is by definition an observable trait. The first step in experimental research is to demonstrate variability of conspicuous consumption within the male members of a population. The reason is obvious: selection processes of whatever type can only act on traits that vary within a population.

One of the major obstacles in the research of human conspicuous consumption is that it is not sufficient to demonstrate variation in conspicuous consumptions. Evolutionary researchers must also demonstrate that the differences in male conspicuous consumption are perceived by female conspecifics. In other words, it must be shown that members of the female sex can discriminate between different rates or intensities of conspicuous consumption. Merely demonstrating that conspicuous consumption is observable is of no use unless it is possible to show that there is variation in this trait and that there are consistent female preferences for a specific part of the variation of conspicuous consumption.

This brings me to my second point: the perception of conspicuous consumption behavior needs to be examined in a reproduction or mating context. For sexual selection, there must be some evidence that conspicuous consumption elicits female choices or psychological responses that are consistent with the possibility of conception. In other words, conspicuous consumption has to be experimentally tested at a time when the perception of the display can lead to successful reproduction. Hence, tests with university students (Griskevicius et al. 2007, Sundie et al. 2011) are useful only if it can be also shown that the same display have a similar effect when females make decisions about conception later in life. Especially in western societies where most experiments take place, it is a common practice for young couples to go through a prolonged period of their relationship where sex is not intended for conception. Female responses to conspicuous consumption during this period may not be informative about sexually selected displays when conception is intended.

4.2 Conspicuous Consumption is Costly

The high-cost criterion is commonly cited among evolutionary researchers. For example, Graskevicious et al (2007) and Meijers and Nelissen (2011) argue that conspicuous consumption is costly to males in terms of resources such as money, time, energy or risk, whereby the costlier the trait the more likely it is to be an honest display.

To qualify as a cost in line with theoretical models, empirical studies must demonstrate that sexual traits have a detrimental effect on the survival component of fitness. Moreover, it must be demonstrated that the costs are differential such that negative effects on fitness are more severe for those males with poor quality traits than for those with better quality traits.

We have seen in section 2 that costs of sexual adaptations are of central importance to the theory of sexual selection and costly signalling theory. In the evolutionary models of Fraja (2009) and Collins, Baer and Weber (2015), costs have been explicitly described as having a detrimental influence on the survival of those that conspicuously consume. If fitness is measured in terms of offspring produced throughout the lifetime of an individual, then to count as an evolutionarily significant cost, an increase of conspicuous consumption has to lead to a decrease in survivability of an organism and an increase in the number of offspring it produces.

In the experimental literature however there is often confusion over what qualifies as a cost. Typically, costs refer to an increase in resource expenditure (Graskevicious et al 2007, Meijers and Nelissen 2011, Sundie et al. 2011). Conspicuous consumption is often associated with the purchase of expensive products that are consumed in public and are used by males to display material wealth. However, it is important to make a distinction between resource expenditure and fitness costs: conspicuous consumption may require a high expenditure of money or time without being costly in terms of reduced lifetime male survivability. Consequently, the claim that conspicuous consumption is costly relies almost exclusively on indirect evidence about resource expenditure while empirical evidence on actual influence on lifetime reproduction success is at best circumstantial.

There is an additional problem with the handicap principle (i.e costly signalling theory) frequently cited by researchers as a mechanism for ensuring the reliability of conspicuous consumption. The costly signalling hypothesis proposes that the costs associated with the production or maintenance of the display is crucial for its reliability. However, recent critiques of costly signalling theory suggest that a decrease in a fitness component incurred by a display trait is not necessary to prevent deception (Grose 2011, Számadó 2011). In other words, a high display cost is not the only mechanism that maintains honesty and cheap or even cost-free displays can be honest.

For example, according to the index hypotheses, the actual fitness costs of honest displays could be zero. A display that is not difficult to produce or maintain can still be honest if there is causal link between a male quality (e.g. material wealth) and the

display (e.g. conspicuous consumption) that makes it difficult for low quality males to produce (Dezecache, Mercier and Scott-Phillips 2013). As long as high quality individuals and not low-quality individuals gain a net fitness benefit from giving a high-quality signal, costs paid for giving a display are arbitrary in maintaining honesty. Cheap or even cost-free mechanisms can also maintain display honesty although they have largely been ignored in the research of conspicuous consumption.

To sum up, in the recent literature fitness costs are confused with expenditure. Costs in terms of resources are suggested to provide evidence for the fitness costliness of conspicuous consumption. In line with theoretical requirements, more care is needed in future empirical studies to distinguish fitness costs from resource expenditure. In addition, evolutionary researchers working on conspicuous consumption must justify their preference of particular evolutionary models over other alternatives that provide evolutionary explanations of why conspicuous consumption is a reliable display.

4.3 Conspicuous Consumption provides Fitness Benefits

The third criterion is that conspicuous consumption must increase the probability that the male signaler will gain some fitness advantage through the display, such as increased ability to attract and copulate desirable mates. This is not easy to experimentally test in human populations, mainly because it is difficult to demonstrate that those that have a display trait produce more offspring through their lifetime. There is not a single study that shows a correlation between male conspicuous consumption and differential reproductive benefits. At best, there is plenty of anecdotal evidence that human females find male conspicuous consumers attractive. In the absence of empirical studies that demonstrate that differences in a conspicuous consumption actually result in differential reproductive success, claims that conspicuous consumption is a sexual adaptation is simply a hypotheses in need for empirical support.

There is an additional and more serious problem in the claim that conspicuous consumption is a sexual adaptation. Empirical studies in evolutionary psychology suggest that human females prefer high status males or those that demonstrate the potential to attain high status (Buss 1989 and 2016, Geary 1998). Using this insight, evolutionary researchers often argue that display traits such as conspicuous consumption are used by human males to maintain or increase their status. This argument finds conspicuous consumption to have an intermediary role: an increase in conspicuous consumption leads to an increase in a male's relative status and to a

decrease in the relative status of other competitors. Females prefer higher status mates and therefore there is an increase in male reproductive success.

What is problematic with this argument is that in this literature it is also assumed that conspicuous consumption has negative or minimal influence on survivorship (Fraja, 2009, Collins Baer and Weber 2015). The pursuit of social status via conspicuous consumption however can provide direct or indirect survival benefits for high status males in addition to mating access. Such benefits include deference, coalitions, aid in childcare or after injury and illness, leniency after a failure to reciprocate. The benefits derived from having status can be cumulative since they are received from multiple conspecifics. Hence the overall impact of high status on fitness is not only an increase in fertility, but also an increase in the viability (or well-being depending on how viability is conceived) of self, spouse, offspring and other kin as well.

4.4 Conspicuous Consumption is an Indicator of Male Quality Traits

By their very construction, evolutionary models of conspicuous consumption treat all male qualities traits as exogenous. Females are assumed to have a fixed preference for whatever male qualities are displayed (Collins Baer and Weber 2015: p.194). Moreover, quality traits are not heritable but are randomly allocated at birth. These assumptions allow tractable models for the theoretical analysis of conspicuous consumption (Fraja 2009, Collins Baer and Weber 2015: p. 203). However, this also creates ambiguity about what kind of attributes males convey by the display trait. Accumulated resources (i.e. wealth), potential to acquire resources, dexterity, altruism, courage, health, status, physical condition, artistic skills, ambition, industriousness, intelligence or (unspecified) genetic qualities are mentioned in this literature as quality attributes indicated by conspicuous consumption. Whether there is a reliable link between a males' conspicuous consumption and these qualities cannot simply be assumed to be true but require experimental and theoretical investigation.

In what follows I will discuss a possible response to the problem of determining particular quality traits. It is possible to argue that females choose mates that conspicuously consume because this is a good indicator of a male's *overall* quality. When it is difficult to determine *specific* quality traits of potential mating partners, a female can use the quality of conspicuous consumption as a proxy of a male's composite ofgenetic or phenotypic qualities. Whether this hypothesis is plausible dependson the environment in which mating interactions are taking place. For example, in small and stable groups, females would not be so much interested in males' conspicuous consumption because they could acquire information about specific qualities of potential mates through gossip or repeated interactions.

To conclude this section, it is doubtful whether conspicuous consumption meets the criteria of costly signalling theory and sexual selection. As we will see in the next section, problems only multiply when we consider that the male display trait and female preference are supposed to have evolved sometime during the Pleistocene.

5 Conspicuous Consumption and Evidence from Hunter-Gatherer Societies

In order for evolutionary researchers to make their argument they must offer reasonable evidence that conspicuous consumption does in fact possess the characteristics presented in the previous section. We have seen that there are reasons to doubt this claim. Suppose however that evolutionary research has experimentally demonstrated that conspicuous consumption is an easily observable, costly trait that confers reproductive success to males. Let us suppose even further that evolutionary research has shown that females have a preference for males' level or quality of conspicuous consumption. A question left to ask is whether these researchers have shown that conspicuous consumption is a sexual adaptation.

Most of the published literature implicitly assumes that to demonstrate that a trait such as conspicuous consumption has the characteristics presented in section 2 is sufficient to be able to conclude that a trait evolved by sexual selection. In current research there is little effort in establishing a plausible relation among the display trait, quality traits and the preference traits by using available information about the ancestral environment. It appears that evolutionary research has lost track of the kind of evidence needed for the empirical support of sexual selection explanations. Even if evolutionary researchers demonstrate that conspicuous consumption has the characteristics presented in the previous section, one has no reason to think that it is likely brought about by sexual selection. At best, evidence about current behavior and current mating success can only suggest that sexual selection acts in the present environment. No matter which traits one chooses to examine, evidence about the ancestral environment is essential to support hypotheses about sexual adaptations.

My objection to this evolutionary explanation of conspicuous consumption is the following: if researchers claim that conspicuous consumption is a sexual adaptation to a particular environment, it is essential to present in detail what that environment is like. Conspicuous consumption needs to be understood according to its unique, time and place in human evolution. If it is argued that evolutionary researchers are aware of the present-centered starting point of their investigation but think that knowledge of how sexual selection is operating in the present allows them to extrapolate back in time, I would agree if they can also demonstrate that the selection conditions have remained constant. In turn, examining whether this assumption holds requires evidence about our ancestral past.

It follows that it is important is to examine the selection conditions in which traits evolved. Fraja (2009) seems to agree and argues that conspicuous consumption *"was hard wired in the brain of early humans prior to their dispersion from Africa, and therefore must have provided evolutionary advantages in the conditions prevailing between one million and 80,000 years ago"* (p.52). In order to get antecedent knowledge about the social organization, mating patterns and food acquisition in the ancestral past, anthropologists and ethnographers typically examine present hunter-gatherer societies not conducive to agriculture (e.g. Ember 1978; Marlowe 2005). These foraging groups might have changed significantly over time but they can shed light on the social environment in which the evolution of human traits has taken place.

What can we infer about our distant ancestors by looking at the few thoroughly studied hunter-gatherer societies of recent times? Ethnographers that study modern foraging groups as a proxy of ancestral hunter-gatherer societies[20] have arrived at some generally accepted conditions about late Pleistocene social life:

a) Population density was low and groups were small with an average size no more than 100 members[21] (Marlowe 2005, Bowles and Gintis 2011);

---

[20] Although there is a great diversity among these groups (Kelly 2007, Henrich and McElreath, 2007), anthropology typically distinguishes between two categories of hunter-gatherer societies (Woodburn 1982): delayed-return and immediate-return. Immediate-return societies are considered a good proxy of life in the human past while more modern delayed-return societies demonstrate among others marked inequalities, social hierarchy and larger group size. As any other categorization, it is not absolute; in some foraging groups, it is possible to find characteristics of both delayed and immediate-return societies. The hunter-gatherer societies examined in the present chapter belong to the immediate-return category. The evidence strongly suggests that the environment in which humans evolved did not include domesticated plants and animals, food surpluses or storing mediums.

[21] If the average camp size is very small, a group will require regular genetic inputs from other camps in order to stay viable. Moreover, the boundaries of hunter-gatherer groups are often fluid and local bands fission or fuse in search of resources (Marlowe, 2005). This, however, does not necessary implies that when males move to other camps they engage in conspicuous consumption to attract mating partners that belong to other bands. I thank KristVaesen for bringing these points to my attention.

b) Groups of hunter-gatherers were highly mobile and nomadic, moving frequently from one territory to another to get food resources necessary for subsistence;
c) There was little difference in resources among individuals because sharing practices were common;

If we accept that hunter-gatherer populations have the above characteristics, there are two important implications for the evolutionary explanation of conspicuous consumption based on costly signaling theory and sexual selection.

First, a display trait such as conspicuous consumption would have little value in small and stable groups either because there is the continuous interaction between the same group members or because private information would have become public through third parties that communicate the quality traits of group members. In long-lasting, stable hunter-gatherer groups, it is more likely that mate choice is based on the accumulation and integration of information over a long period of time through repeated encounters or over the course of development[22].

More generally, females have an incentive to respond to a display trait such as conspicuous consumption when they face information asymmetry about male private quality traits such as accumulated resources, status, skills related to the acquisition of food or the construction of privately owned objects. Precisely because there is female uncertainty about these traits, males are motivated to conspicuously consume and females value this display trait. In the context of foraging societies, demographic factors such as small group size mitigate information asymmetry and the value of a display trait such as conspicuous consumption.

Second, these nomadic hunter-gatherer groups strive to meet daily subsistence requirements without devices that allow the storage of a surplus in resources. They often emphasize sharing in the distribution of resources and reciprocation to endure harsh environmental conditions. The implication of anthropological research is that even if there is the inequality in hunting abilities, there is near equality in lifetime consumption of food resources. When the prey is not consumed near the place where it is captured, specific sharing rules prescribe that it is distributed not only among the nuclear family but also among other member of the group. In contrast to modern societies, in these hunter-gatherer groups inequality in the production of resources (i.e. capturing game) does not necassarily result to inequalities in accumulated resources or consumption of resources. Because storable and therefore inherited wealth (i.e. accumulated resources) is nonexistent,

---

[22] Hunter-gatherers appear to promote play from a very young age quite deliberately in order for children to acquire the skills, knowledge, and values of their group (Gray, 2014). From a very young age and until they reach reproductive maturity, females have ample opportunity to evaluate the skills of potential mates that will become successful adults in their group.

conspicuous consumption that displays wealth cannot be the used as source of value in a mating context.

Since sharing norms and the nomadic lifestyle prevent saving and accumulation of resources, it leaves the production and possession of skilfully made objects (i.e. tools, weapons, ornaments) to act as a signalling medium directed to potential mates. Collins, Baer, and Weber (2015: p.191) and Fraja (2009; p.62) suggest that the work of Kohn and Mithen (1999) on the elaborate morphology of Palaeolithic Acheulian handaxes supports the hypothesis that conspicuous consumption is a sexual adaptation. According to Kohn and Mithen (1999), impractically large hand axes made from about 1.4 Myr ago to around 100,000 years ago were symbols of status, built and displayed by males to attract potential mates.

However, providing an evolutionary explanation of why these hand axes have been built is one of the most controversial research topics in Palaeolithic archaeology. After decades of research, it is still difficult for archaeologists to establish the purpose of this important tool found in many different locations around the world. Nowell and Chang (2009) review the available ethological and archaeological evidence and provide a detailed critique of the Kohn and Mithen's explanation for handaxe morphology. In accord with the analysis in this chapter they conclude that

"the evidence strongly suggests that variation in handaxe morphology is governed by a number of complex factors that differ in influence over time and space, rather than a single overarching mechanism such as sexual selection. As with most questions of interest in Palaeolithic archaeology, we are better served by context-specific, historically situated explanations rather than monocausal scenarios, particularly when our desire is to understand the form and function of an artefact such as the handaxe, which is found in the record, in one form or another, for more than 1.5 million years" (p. 84).

My suggestion is that conspicuous consumption is a kind of behavior undertaken in a social context that already includes aspects such as rapid population growth, high population densities, technological innovations, political institutions and formalized hierarchies based on inheritance of wealth. Human societies grew in complexity due to labour specialization and notable inequalities in the production and the consumption of resources began to emerge. Moreover, production lost its privileged position and consumption became the mean through which individuals present themselves to others.

This suggestion is explicit in Veblen's work. In the pages of the first chapters of the Theory of the Leisure Class, Veblen (1899) provides the social and historical background for the evolution of consumption phenomena. In contrast to contemporary research on conspicuous consumption, Veblen's analysis is based on anthropological work of his time which allowed him to distinguish modern market societies from traditional societies that lived at edge of subsistence struggling to

acquire the basic means for survival.[23] Without references to particular historical periods, Veblen positioned conspicuous consumption within a particular social and cultural context and examined the acquisition and display of commodities as means of securing social differentiation and status in contemporary societies that have developed property rights and accumulated monetary and material resources.

Apart from distinguishing between modern and ancestral environment, Veblen's second important contribution is that he analyzed evolved psychological aspects of human conspicuous consumption. He argued that "irreducible elements of human nature", what he called "instincts", might have a significant influence on evolutionary processes (1914, p.3). According to Richerson and Christiansen (2013), Veblen's work on human predispositions "clearly foreshadows the notion of epigenetic rules, cultural selection, and biased transmission that figure in the late twentieth century revival of Darwinian theories of cultural evolution" (p.7). In the section that follows I will expand on this insight to provide an alternative explanation of conspicuous consumption.

6. The Cultural Transmission of Conspicuous Consumption

In the previous sections I argued that the evolutionary explanation of conspicuous consumption based on sexual selection faces three conceptual and empirical problems. First, it is not clear what kind of traits researchers have in mind by conspicuous consumption. Second, conspicuous consumption does not conform to theconditions for the reliability of costly signalling traits. Finally, the available ethnographic evidence from hunter-gatherer societies casts doubt whether conspicuous consumption evolved as a mating strategy. In relation to these problems my suggestion was that conspicuous consumption is better conceived as a type of overt behavior that evolved in the context of modern market economies. Such a reading is in accord with cultural evolutionary theory. An open question is what motivates individuals to engage in conspicuous consumption? Current evolutionary research on conspicuous consumption has neglected the importance of social learning mechanisms that motivate conspicuous consumption.

Before answering this question, I will present some key aspects of a framework provided cultural evolution theorists (Boyd and Richerson 1985, Richerson and Boyd 2005). I will then explore more in depth areas of research where cultural evolution can provide an original contribution.

---

[23] He summarizes his views in the following way: *"They are small groups and of a simple (archaic) structure; they are commonly peaceable and sedentary; they are poor; and the individual ownership is not a dominant feature of their economic system."* (Veblen 1899: p.8)

According to Richerson and Boyd (2005), the main reason that modern society is different from Paleolithic society is due to extensive and cumulative cultural evolution. From an evolutionary perspective, "culture" refers to ideas, skills, beliefs, habits, behaviors and values transmitted from one individual to another via imitation, teaching and other forms of social learning (Boyd and Richerson, 1985, Henrich and McElreath 2003, Richerson and Boyd 2005). This definition of culture emphasizes the importance of cognitive abilities to acquire behaviors, ideas etc. by observing or interacting with others. For example, a child that reaches puberty may start to imitate cultural traits of her teachers or her friends and abandon traits acquired from her parents. This can be the result of peer pressure or personal experience that made her realize that parent imitation is not useful anymore.

In relation to the arguments in the previous sections, an important advantage of cultural evolution theory is that it is empirically accessible to historical observation in modern societies, whereas sexual selection must be studied using evidence from the ancestral environment. Thus, it is easier to identify and subsequently empirically support what motivates conspicuous consumption. A second related advantage is that it is possible to impose a taxonomy that collects the diverse concrete reasons that produce conspicuous consumption into classes with similar features. The processes of selectively acquiring particular traits are attributed to different forms of "biased transmission". A final advantage is that it is possible to build evolutionary models that link multiple individual social learning biases to population-level dynamics.

Given that conspicuous consumption is taking place in modern market societies, cultural evolution theory helps us understand how learning biases motivate conspicuous consumption. Cultural evolution work relevant to understanding conspicuous consumption can focus on two different areas: (i) how evolved social learning biases motivate conspicuous consumption generate and (ii) how these biases maintain different frequencies of conspicuous consumption in a population.

The process of selectively acquiring variants is attributed to "biased transmission" (Henrich and McElreath 2003). Cultural evolution theory typically distinguishes the following types of learning biases:

- Content bias (or "direct" bias) is similar to economists' cost–benefit analysis (often called asocial learning). An individual selects a cultural trait based on its perceived value.
- Frequency-dependent biases refer to the relationship between the frequency of a trait and its probability of adoption. Individuals may put higher or lower weight to more frequent traits (conformity bias and anti-conformity bias).
- Context bias (or "indirect" bias) refers to individuals acquiring traits based on the characteristics of an individual (or "model") who exhibits it. Social learners use a cue such as success in a particular domain to determine whether a particular person is an appropriate model.

Typically research focuses on two psychological social learning biases that influence the transmission of ideas and behaviors across multiple domains: conformist and prestige bias. Conformist bias is particularly important because it homogenizes groups by maintaining or reducing within-group differences relative to variation between groups, driving cultural group selection and the evolution of cultural traits (Henrich 2012, ChudekMuthukrishna and Henrich 2015). Non-conformist or anti-conformist bias is an alternative frequency-dependent bias in the acquisition of a cultural trait that increases within-group diversity (Efferson et al. 2008).

Another important component of human cognition is prestige bias.[24] Prestige refers to social recognition of particular individuals in a group for having some valued characteristics. It is suggested that there is a strong propensity to preferentially copy the ideas, behaviors, values and opinions of prestigious or successful individuals. Prestige bias is important when there is limited information about who are the most successful individuals in a group. In this case, individuals might use the different displays of others as short-cuts of whom to copy. This bias can lead to the spread of maladaptive traits that correlate with ideas, practices and behaviors that promote reproductive success (Boyd and Richerson 1985).

In line with the above categorization of biases, it is possible to classify different motivational factors involved in conspicuous consumption.

First, according to content bias, individuals adopt a consumption behavior by observing and comparing its main characteristics (color, design) with other alternatives. Individuals may consume specific goods simply because they think that some of their features are better than others goods. Even if some objects are acquired and consumed primarily for their display value, most goods have also use value. Consumers might acquire conspicuous goods because of their speed, design or comfort. To use an example from the literature, a consumer may buy a Porsche car instead of a VW car because it allows her to move fast from one location to another with relative safety.

Second, according to context bias, individuals have the propensity to adopt a consumption pattern that is most common or rare in the population or a group. Although future empirical research might test which of the two context biases – conformity or anti-conformity - is more potent in the case of conspicuous consumption, both provide insights into human consumption behavior.

Individuals often engage in conspicuous consumption to conform to the majority of the group or the population they are members. Expenditure to acquire expensive goods to communicate a conforming image is often necessary to maintain social standing or gain acceptance by other group members. This propensity to conform to the expectations of groups with whom an individual is associated with is reflected in

---

[24] For a thorough analysis of prestige bias together with evidence from ethnography and social psychology, see Henrich and Gil-White (2001).

the common saying "keeping up with the Joneses." Psychological research has demonstrated that people tend to conform to the majority of the group they are member when forming ideas, behaviors and values. In cases of conspicuous consumption, this motivation often is more significant and pronounced. Hence, the intuitive suggestion of cultural evolution theory is that in addition to relative social standing, similar positions in a social hierarchy influence the adoption of particular consumption patterns.

Conversely, consumers might have the propensity to differ from the consumption behaviors that are most common in the population or a group[25]. One key characteristic of conspicuous consumption is exclusivity - consumers desire to be perceived as different from the members of the group they belong or the population in general. Hence, goods considered scarce or unique appear to have higher value for consumers. Moreover, the perceived exclusivity of particular goods decreases when more consumers purchase and demonstrate them in public. To return to the Porsche example, although there is a long waiting list for Porsche cars, in order to maintain their sales, Porsche limits the quantity of cars produced by their factories to a few thousand vehicles per year. This is also the case with goods such as watches, jewellery, perfumes or collector's items such as stamps. Producers of consumption goods highlight their exclusivity since many consumers would find a Porsche less valuable if it is widely available. An insight of cultural evolution theory compared to sexual selection is that rather than to display their wealth, social status or particular quality traits to potential mates, consumers would often choose to buy and display expensive goods just for the sake of being different from other consumers.

Third, consumers may have a propensity to acquire a specific consumption behavior based on the observable attributes of a model individual that exhibits particular values, ideas and behavior that are considered important in a group. Veblen (1899) was one of the first theorists who argued that consumption is actually a process of emulation and goods function as markers of model individuals. According to this idea, individuals emulate the consumption patterns of those that are perceived as higher in the hierarchy of a group or the population. However, in modern times, emulation is a round or two-directional process: it can be top-down or bottom-up. According to Trigg (2001), Bourdieu (1984, 1990) provides a contemporary development of the theory of conspicuous consumption. Although consumers typically emulate the consumption pattern of prosperous or prestigious individuals, Bourdieu's hypothesis is that emulation can also be bottom-up: those with a higher position in the social hierarchy might emulate the behavior and consumption patterns of those in a lower position. Individuals with a higher status may increase their reliance on social learning and copy those who they perceive as having lower statues because they see them as superior sources of information.

---

[25] In economics this was first described by Leibenstein (1950) as "snob effect."

The psychological biases described in the previous paragraphs can be understood as individual level mechanisms or as population-level properties that affect the rate of transmission of traits within or between groups. Evolutionary researchers can take inspiration from both population genetics and social psychology to build models of conspicuous consumption, emphasizing particular social learning biases. Surprisingly, there has been very little formal modelling of how social learning influences the evolution of display traits such as conspicuous consumption. As far as I am aware of there is only one model related to indirect and direct bias in the evolution of conspicuous consumption (Boyd and Richerson, 1985: ch.8, for a discussion see Cordes, 2009). Boyd and Richerson demonstrate that a runaway process can lead a display trait such as conspicuous consumption to take exaggerated forms. Unfortunately, there is no formal work on the role of conformity and anti-conformity transmission in the emergence and stability of conspicuous consumption although these psychological propensities can be easily modeled as frequency-dependent selection bias.

To conclude, cultural evolution theory provides a framework that allows researchers to identify and classify the psychological motives of conspicuous consumers. In contrast to an explanation based on sexual selection that focuses on displays to attract potential mates, I suggest that our psychological learning biases have a crucial motivating role in consumption decisions, including conspicuous consumption.

Evolutionary researchers might raise the following objection: the analysis in this section refers to proximate behavior-generating mechanisms, not evolutionary causes such as sexual selection. My response is based on Henrich and McElreath's (2007) work on the evolution of social learning biases:

- Genetic evolution gives rise to psychological adaptations that allow the acquisition of information such as ideas, beliefs or strategies of other individuals.
- These adaptations allow the evolution of the cultural system of inheritance independent from genetic inheritance. Humans become gradually reliant on social learning and cultural evolution could arise.
- Cultural evolution processes can change the selection conditions on which natural selection operates.

Evolutionary approaches to culture suggest that our species' social learning biases – the psychological mechanisms that underlie the evolution of cultural traits such as conspicuous consumption – are genetically evolved adaptations in ancestral environments. It is social learning mechanisms that evolved to enhance fitness while conspicuous consumption taps into psychologies already in place. Put in a different way, social learning biases arose some time during our evolutionary history because they were favored by natural selection and have remained an important feature of human psychology and social interactions since that time. Hence the question that

needs to be addressed is whether genetic evolution gave rise to psychological adaptations that allow the cultural evolution of conspicuous consumption.

In the last 30 years, evolutionary theorists have produced a large body of evolution models that examine the dynamics of learning biases and have tested aspects of these models. Within the framework of cultural evolution, formal work explores the conditions of the environments under which natural selection will favor particular social learning strategies and biases (e.g. Boyd and Richerson, 1985, Henrich and Boyd, 1998, Rendell et al 2009, Nakahashi et al 2012). In particular, theoretical research predicts under what conditions social learning biases or individual learning evolve and whether it is evolutionarily advantageous to apply multiple social learning biases. It is argued that if environments did not change rapidly or if the rate of migration between groups was low, conformist or prestige bias would have been favored over individual learning. The core idea is that in rapidly changing environments, social transmission will be of little benefit since the information accumulated from past generations would be quickly outdated. In such environments, asocial learners that experiment with alternative traits will perform better since they mostly acquire information directly from their current environment.

7. Conclusion

This chapter explores the evolution of what is commonly referred to as conspicuous consumption in ancestral and contemporary societies. With respect to sexual selection this chapter leads to the following conclusions: a) the study of sexual selection in human primates needs more critical examination of the assumptions made and the data gathered and b) the available data demonstrate that our ancestors lived in small, stable, long-lasting egalitarian groups and therefore it is doubtful that conspicuous consumption is a sexual adaptation. Females were likely to accumulate information about potential mates from multiple sources over a long time frame.

Although our psychological capacities may have a long history in the human lineage, conspicuous consumption appears relatively late in our evolutionary history. After examining the hypothesis that conspicuous consumption is a sexual adaptation, I suggest that of cultural evolution theory can be essential for elucidating both the evolved aspects of our psychology and the ecological and cultural factors that precipitate and maintain conspicuous consumption in the population.

Explaining the complex phenomenon of conspicuous consumption is still not complete. This is evident in the continued interest of researchers. The exact nature of human motives involved in conspicuous consumption has not been fully investigated and understanding the psychology of conspicuous consumption behavior remains a largely unresolved issue. Future studies might examine the

psychological and evolutionary side of the issue and evaluate possible antecedents of conspicuous consumption.

In principle, building evolutionary models that combine economic, biological and cultural factors is a shared objective among evolutionary scientists. In the case of conspicuous consumption however, little work has been done to actually provide evidence about the ancestral environment and there is little explicit focus on how details about human psychological capacities influence adaptive behavior. Often in the literature, evolutionary researchers apply models that take inspiration from animal models of behavior. Cultural evolution's commitment to develop evolutionary models that are informed by cognitive science, evolutionary theory and evidence about the selection environment might provide novel insights into the evolution of conspicuous consumption.

References

Aoki K, Nakahashi W (2008) Evolution of learning in subdivided populations that occupy environmentally heterogeneous sites. TheorPopul Biol74: 356–368 doi: 10.1016/j.tpb.2008.09.006

Bliege Bird R, & Smith EA (2005) Signaling Theory, Strategic Interaction, and Symbolic Capital. CurrAnthropol46: 221–248 doi: 10.1086/427115

Bourdieu P (1984) Distinction: a social critique of the judgement of taste: Pierre Bourdieu. Harvard University Press, Cambridge MA

Bourdieu P (1990) The logic of practice. Stanford University Press. Stanford CA

Bowles S, Gintis H (2011) A cooperative species: human reciprocity and its evolution. Princeton University Press, Princeton

Boyd R, Richerson PJ (1985) Culture and the evolutionary process. University of Chicago Press, Chicago

Buss D (1989) Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. Behav Brain Sci12: 1–14 doi: 10.1017/s0140525x00023992

Buss D (2016) The evolution of desire: strategies of human mating. Basic Books, New York

Chudek M, Muthukrishna M, Henrich J. (2016). Cultural evolution. In: Buss D (eds) The handbook of evolutionary psychology: Integrations. John Wiley & Sons, New Jersey, pp 749–769. doi: 10.1002/9781119125563.evpsych230

Collins J, Baer B, Weber EJ (2015) Sexual selection, conspicuous consumption and economic growth. J Bioecon17: 189–206. doi: 10.1007/s10818-015-9200-9

Cordes C (2009) The Role of Biology and Culture in Veblenian Consumption Dynamics. J Econ Issues, 43: 115–142. doi: 10.2753/jei0021-3624430106

Darwin, C. (1871/1901) The descent of man and selection in relation to sex. J. Murray, London

Dezecache G, Mercier H, Scott-Phillips TC (2013) An evolutionary approach to emotional communication. J Pragmati59: 221–233. doi: 10.1016/j.pragma.2013.06.007

Efferson C, Lalive R, Richerson P, Mcelreath R, Lubell M (2008) Conformists and mavericks: the empirics of frequency-dependent cultural transmission. Evol and Hum Behav, 29: 56–64. doi: 10.1016/j.evolhumbehav.2007.08.003

Ember CR (1978) Myths about Hunter-Gatherers. Ethnology17: 439. doi: 10.2307/3773193

Fraja GD (2009) The origin of utility: Sexual selection and conspicuous consumption. J Econ Behav Organ, 72: 51–69. doi: 10.1016/j.jebo.2009.05.019

Geary DC (1998) Male, female: The evolution of human sex differences. American Psychological Association, Washington DC

Grafen A (1990a) Biological signals as handicaps. J TheorBiol, 144: 517–546. doi: 10.1016/s0022-5193(05)80088-8

Grafen A (1990b) Sexual selection unhandicapped by the fisher process. J Theor Biol, 144: 473–516. doi: 10.1016/s0022-5193(05)80087-6

Gray P (2014) Play Theory of Hunter-Gatherer Egalitarianism. Ancestral Landscapes in Human Evolution, 192–215. doi: 10.1093/acprof:oso/9780199964253.003.0014

Griskevicius V, Tybur JM, Sundie JM, Cialdini RB, Miller GF, Kenrick DT (2007) Blatant benevolence and conspicuous consumption: When romantic motives elicit strategic costly signals. J Pers Soc Psychol, 93: 85–102. doi: 10.1037/0022-3514.93.1.85

Grose J (2011) Modelling and the fall and rise of the handicap principle. Biol Philos26: 677–696. doi: 10.1007/s10539-011-9275-1

Henrich J, Boyd R (1998) The Evolution of Conformist Transmission and the Emergence of Between-Group Differences. Evol Hum Behav19: 215–241. doi: 10.1016/s1090-5138(98)00018-x

Henrich J, Gil-White FJ (2001) The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. Evol and Hum Behav22: 165–196. doi: 10.1016/s1090-5138(00)00071-4

Henrich J, McElreath R (2003) The evolution of cultural evolution. EvolAnthropol Issues News Rev12: 123–135. doi: 10.1002/evan.10110

Henrich J, McElreath R (2007) Dual-inheritance theory: the evolution of human cultural capacities and cultural evolution. Oxford University Pressdoi: 10.1093/oxfordhb/9780198568308.013.0038

Kelly RL (2007) The foraging spectrum: diversity in hunter-gatherer lifeways. Percheron Press, New York

Kohn M, Mithen S (1999). Handaxes: products of sexual selection? Antiquity73: 518–526 doi: 10.1017/s0003598x00065078

Kurland JA (2001) The Mating Mind: How Sexual Choice Shaped the Evolution of the Human Mind. Am Anthropol103: 1196–1198. doi: 10.1525/aa.2001.103.4.1196

Lee RB (2009) The!Kung San: men, women, and work in a foraging society. Cambridge University Press, Cambridge

Leibenstein H (1950) Bandwagon, Snob, and Veblen Effects in the Theory of Consumers Demand. Q J of Econ64: 183. doi: 10.2307/1882692

Marlowe FW (2005) Hunter-gatherers and human evolution. EvolAnthropol Issues News Rev14: 54–67. doi: 10.1002/evan.20046

Mcelreath R, Henrich J (2007) Modelling cultural evolution. Oxford University Press. doi: 10.1093/oxfordhb/9780198568308.013.0039

Miller G (2001) The mating mind: how sexual choice shaped the evolution of human nature. Vintage, London

Miller G (2009) Spent sex, evolution, and consumer behavior. Penguin Group Usa, New York

Nakahashi W, Wakano JY, Henrich J (2012) Adaptive Social Learning Strategies in Temporally and Spatially Varying Environments. Hum Nat23: 386–418. doi: 10.1007/s12110-012-9151-y

Nelissen RM, Meijers MH (2011) Social benefits of luxury brands as costly signals of wealth and status. Evol Hum Behav32: 343–355. doi: 10.1016/j.evolhumbehav.2010.12.002

Nowell, A, Chang ML (2009) The case against sexual selection as an explanation of handaxe morphology. Paleoanthropology: 77–88. Retrieved from http://www.paleoanthro.org/static/journal/content/PA20090077.pdf

Rendell L, Fogarty L, Hoppitt WJE, Morgan TJ, Webster MM, Laland KN (2011) Cognitive culture: theoretical and empirical insights into social learning strategies. Trends in Cognit Sci15: 68–76. doi: 10.1016/j.tics.2010.12.002

Richerson PJ, Boyd R (2005) Not by genes alone: how culture transformed human evolution. The University of Chicago Press. Chicago

Richerson PJ, Christiansen MH (2013) Cultural evolution: society, technology, language and religion. MIT Press, Cambridge, MA

Saad G (2014) The evolutionary bases of consumption. Psychology Press, New York

Smith JM, Harper D (2003) Animal signals. Oxford University Press, Oxford

Sundie JM, Kenrick DT, Griskevicius V, Tybur JM, Vohs KD, Beal DJ (2011) Peacocks, Porsches, and Thorstein Veblen: Conspicuous consumption as a sexual signaling system. J Pers Soc Psychol100: 664–680. doi: 10.1037/a0021669

Számadó S (2011) The cost of honesty and the fallacy of the handicap principle. AnimBehav81: 3–10. doi: 10.1016/j.anbehav.2010.08.022

Trigg A. B. (2001) Veblen, Bourdieu, and Conspicuous Consumption. J Econ Issues, 35: 99–115. doi: 10.1080/00213624.2001.11506342

Trivers R (1972) Parental investment and sexual selection. Biological Laboratories, Harvard University, Cambridge, MA

Veblen T (1914) The instinct of workmanship and the state of the industrial arts. W. W. Norton And Company, New York

Veblen T (1970) The theory of the leisure class: an economic study of institutions. London, Allen And Unwin

Wakano JY, Aoki K (2007) Do social learning and conformist bias coevolve? Henrich and Boyd revisited. TheorPopul Biol72: 504–512. doi: 10.1016/j.tpb.2007.04.003

Witt U (2016) The evolution of consumption and its welfare effects. J Evol Econ 27: 273-293. doi: 10.1007/s00191-016-0459-3.

Woodburn J (1982) Egalitarian Societies. Man, 17(3), 431. doi: 10.2307/2801707

Zahavi A (1975) Mate selection—A selection for a handicap. J Theor Biol53: 205–214. doi: 10.1016/0022-5193(75)90111-3

Chapter 4:The Red-Beard Evolutionary Explanation of Human Sociality[26]

1. Introduction

Emotion-based communication is a key feature of our daily life, with human interactions being replete with many forms of verbal and non-verbal emotional expressions. Verbal exchanges include speech intonation and the use of language while non-verbal interactions consist of gestures, bodily postures, facial expressions, blushing or perspiration used to convey emotional states. Both verbal and non-verbal expressions of emotions appear to facilitate communication in the social world by providing quick information to others.

Non-verbal displays of emotions have been extensively studied by disciplines such as psychology, neuroscience, economics and biology. Darwinian evolution is widely acknowledged as a key for understanding human emotions. Since Charles Darwin's classic book "The Expression of the Emotions in Man and Animals" (Darwin 1872), evolutionary research on emotions was mostly concerned with identifying and characterizing non-verbal manifestations of discrete emotional states (e.g. Ekman, 1993; Izard, 2010). Joy, fear, surprise, anger, sadness and disgust[27] are considered evolved human characteristics expressed by particular facial expressions across human groups.

More recently, evolutionary theorists systematically link evolutionary explanations of human non-verbal emotional expressions to animal signalling theory (e.g. Dezecache et al 2013; McCullough and Reed, 2016). Moreover, social evolution theory examines phenotypic markers that facilitate assortment and the evolution of social behaviors (e.g. Frank, 1987; 1988; Nesse, 2001; Riolo, Cohen, & Axelrod, 2001). This work has incited a recent trend in empirical research that investigates whether smiles (e.g. Reed, Zeglen, & Schmidt, 2012; Centorrino, Djemai, Hpfensitz, Milinki, & Seabright, 2015), tears (e.g. Gračanin, Bylsma, &Vingerhoets, 2018; Reed, Matari, Wu, and Janaswamy, 2019), blushing faces (e.g. Dijk, Koenig, Ketelaar, & De

[26] A paper based on this chapter is currently under peer review.

[27] Different theorists have different lists of basic emotions. For a recent review, see Keltner et al. (2019).

Jong, 2011) and expressions of anger (e.g. Reed, DeScioli, & Pinker, 2014) are associated with social behaviors. Philosophical research in evolutionary ethics often relies on Robert Frank's older theorizing on social emotions like love, anger, sympathy or jealousy and social behaviors (e.g. Joyce 2006; James, 2011; Heath and Rioux, 2018).

Following this line of work, I discuss recent attempts to explain the evolution of emotional markers and social behaviors. Despite the diversity of evolutionary models in social evolution theory and animal signalling theory, even a cursory look at this literature reveals that a certain issue arises again and again in different contexts. At the theoretical level, many of these evolutionary studies rely on a simple but controversial green-beard complex to explain the evolution of altruistic cooperation (e.g. Gardner and West 2009; West and Gardner, 2010; Cohen, 2012; Handfield, Thrasher & García, 2018). Evolutionary theorists have long emphasized that the green-beard explanation of social behaviors are inherently unstable or short lived. A correlation between marker and social behavior can be disrupted by the spread of a mutant gene, which adopts the phenotypic marker of cooperative individuals and elicit a favorable behavioral response.

In what follows I will not review or evaluate empirical research and experiments that test hypotheses about different emotional expressions[28]. It is also beyond the scope of this paper to present different theories in evolutionary ethics. Instead, I focus on Robert Frank's account (Frank, 1987, 1988, 2005) and the arguments raised against it (e.g. Gardner & West, 2009; Heath &Rioux, 2018). My goal is to strengthen the case for the value of Frank's theoretical work in explaining the evolution of human sociality.

In particular, I argue that recent well put criticism against Frank's account of social behavior does not render his ideas useless. Using resources found in Frank's work, I distinguish between the red-beard evolutionary explanation that targets proximate causes such as emotions that govern behavior and phenotypic markers and the green-beard explanation provided by Robert Frank's evolutionary model that focuses on behavior and fitness outcomes. Once we distinguish between these two different kinds of evolutionary explanation, it transpires that we do not need to abandon the idea of evolutionary stable communication between individuals based on emotions. In the red-beard explanation there a causal link between human emotions, social behavior and phenotypic markers and this makes a substantial

[28]It is an open empirical question which emotions are associated with social behaviors and distinct external markers or whether there are emotional expressions that can be recognized with accuracy higher than random guessing. What complicates empirical investigation is that a given emotion may be associated with a variety of expressions, while various emotions may be related to the same expression. More experimental research is required to demonstrate if there are some emotions that are exclusively connected to few markers and particular social behaviors.

difference in the way one addresses the stability problem. Social emotions are a common cause of a social behavior and a phenotypic marker and therefore cooperative behavior cannot be suppressed without also changing the marker. Hence, the red-beard explanation is more plausible than it might appear at first glance.

The structure of the paper is as follows. In the next section I present the green-beard explanation of cooperation. Section 3 examines conceptual issues related to Robert Frank's evolutionary model. Section 4 distinguishes between the red-beard and the green-beard explanations of cooperation and provides a solution to the stability problem. Section 5 concludes.

2. The Inclusive Fitness Analysis of Human Sociality

In this section I examine ultimate explanations of social behaviors that rely on the green-beard complex. The advantage of doing so is that it allows us to compare explanations as they are provided by different theoretical accounts and focus on what makes evolutionary explanations in the social sciences different in kind.

Social evolution theory is based on the ultimate-proximate distinction (Mayr, 1961). According to Vromen (2017), there are at least three ways to view Mayr's distinction: First, there is a distinction between evolutionary causes (processes such as natural selection) and proximate behavior-generating causes (e.g. psychological, neurological) inside organisms. Second, there is a distinction between distant and nearby parts in the causal chain that lead up to a behavior. Finally, there is a distinction between different explanatory projects: ultimate explanations typically answer why questions (why did social behavior evolve?) while proximate explanations typically answer how questions (how is social behavior produced?).

Based on this insight, Vromen explicates the conventional view to answering ultimate why-questions in social evolution theory. Ultimate explanations examine why behaviors evolved while proximate explanations how they work. For example, if a phenotypic marker is explained in terms of intentional decision-making and behavior-generating motives, researchers provide a proximate explanation[29] . If the same phenotypic marker is explained in terms of the dynamic of natural selection and how, on average, it has positive fitness consequences they provide an ultimate explanation. It is generally accepted that ultimate and proximate explanations complement each other and both are required for a complete account of social behavior.

---

[29]Proximate mechanisms can also refer to neural activity, cognitive mechanisms that regulate social learning or mental states related to beliefs and desires.

In social evolution theory behaviors are typically classified based on different combinations of their positive or negative effect on the actor and others' fitness (e.g. Hamilton, 1964, Gardner et al 2007; Bshary and Burgmüller, 2007). In particular, a behavior that is beneficial to the actor and is costly to the recipient is selfish. Altruistic behavior is costly to the actor and beneficial to the recipient while mutually beneficial behavior is beneficial both to the actor and the recipient's fitness. Finally, a behavior that is costly in terms of direct fitness to both the actor and the recipient is spiteful. Mutually beneficial and altruistic behaviors are often referred as cooperative behaviors.

The challenge in social evolution theory is to explain why social behaviors like altruism could have evolved despite being costly to the organisms that perform them. This puzzle has received a great deal of attention and generated a voluminous body of literature. The solution relies on the idea that the evolution of social behaviors is driven by effects on a focal individual's own direct fitness and effects on the fitness of individuals that have the same gene(s) (i.e. indirect fitness). Based on this distinction, ultimate explanations of altruism and mutually beneficial behaviors are classified into two broad categories: Direct and indirect ultimate explanations. In direct fitness explanations social behavior evolves because of an increase in the direct fitness of the actor. For example, in repeated interactions (Trivers, 1971; Axelrod and Hamilton, 1981) or collaboration (Sterelny, 2016) the actor might gain a delayed direct fitness benefit that is higher than the immediate fitness cost. By contrast, indirect fitness explanations explore the presence of genetic relatedness and indirect fitness benefits. Social behavior evolves because indirect fitness is higher than the direct fitness cost for the actor.

One can further distinguish indirect ultimate explanations of social behavior based on how fitness benefits arise in interactions between individuals that share the same gene(s). There are explanations that rely on genealogical kinship and green beard explanations based on a linkage or pleiotropy that allows the actor to gain an indirect fitness benefit (Hamilton, 1964; Dawkins, 1976). In what follows I focus on the green-beard explanation of cooperation that is relevant to the discussion of emotional markers in the sections that follow.

In the original formulation of the green-beard explanation, Hamilton (1964) considered a specific pleiotropy, a "supergene affecting (a) some perceptible feature of the organism, (b) the perception of that feature, and (c) the social response consequent upon what was perceived" (Hamilton, 1964, pp.25). The supergene hypothesis is explicit about the underlying genetics: one gene influences a set of phenotypic traits. It simultaneously encodes for a social behavioral response, a conspicuous phenotypic marker and the capacity to recognize this feature of an organism.

Dawkins analyzed Hamilton's explanation and coined the widely used term "green-beard" to describe Hamilton's perceptible feature of the "supergene". A single gene

might produce a distinct recognizable marker (i.e. the green-beard)[30] and a tendency to behave in a "nice way" to those that carry that trait (Dawkins, 1976, pp.96). Although it is often unclear what kind of social behavior is part of the green-beard complex, it is typically assumed that the green-beard explanation introduced by Hamilton and Dawkins refers to altruistic behavior (Dawkins, 1982, pp.145). Bearers of a phenotypic marker (i.e. the green-beard) incur a fitness cost to provide direct fitness benefits to those that also have a green-beard. This type of social behavior is thus personally costly to that particular altruist in direct fitness terms, but linkage or pleiotropy result to an increase in indirect fitness and allows altruistic behavior to evolve.

More recent formulations of the greenbeard explanation refer to cases that involve distinct but tightly linked variant gene forms (i.e. alleles) that encode for the visible phenotypic markers and altruistic behavior (e.g. Gardner and West, 2009; West and Gardner, 2010; Biernaskie et al., 2011). Altruistic behavior evolves because of the genetic relatedness between donor and recipient due to linkage within the genome. Genetic linkage typically occurs when a set of variant gene forms are transmitted across generations as a whole because they are close to one another on a chromosome (i.e. the thread like arrangement of DNA in a cell's nucleus). It results to indirect fitness benefits that are higher than direct fitness costs and allows altruism to evolve in a population.

There is an important reason to doubt the evolutionary stability of altruistic behavior that relies on a green-beard complex. Consider a population of altruists that have a phenotypic marker to convey information about their tendency to provide fitness benefits to others. This population can be invaded by non-altruists that bear the marker but do not provide fitness benefits to the altruist. Before long, altruists would stop displaying the marker and the interaction would collapse.

In more detail, an ultimate explanation of altruism that relies on green-beard complex is vulnerable to a modifier gene (i.e. "falsebeard" gene) that emerges in the genome via mutation or recombination and subverts or replaces altruistic behavior but retains the associated phenotypic marker that attracts fitness benefits from altruists. These modifier genes are favoured by selection because they result to higher fitness benefits and altruism is expected to have a short existence over evolutionary time.

This argument poses a challenge to ultimate explanations of cooperation that rely on emotional markers. A common research strategy is to apply insights from social evolution theory to re-examine explanations provided by evolutionary models in the

---

[30]In line with the literature and in order to keep the analysis simple, I assume that recognition capacities are neutral in terms of fitness and do not have an influence on stability. Note, however, that in his evolutionary model Frank assumes that cooperators have to pay a fitness cost to recognize other cooperators.

social sciences (e.g. Gardner et al. 2007; Vromen, 2017). Several evolutionary models attempt to explain human cooperation based on the presence of emotion-based phenotypic markers (e.g. Frank 1987; 1988; Owren and Bachorowski, 2006). Using inclusive fitness analysis, it is argued that these ultimate explanation would not be evolutionarily stable (Gardner and West, 2009; West and Gardner, 2010; Heath and Rioux, 2018).

As the idea of the evolution of emotion-based social behaviors is worked out extensively by Robert Frank, the following discussion focuses on his views. In the following section I examine the structure and results of Robert Frank's evolutionary model and the explanation it provides. Frank (1987) applies tools from evolutionary game theory to examine social behaviors and non-verbal markers. His game theoretic explanation is re-analyzed by taking into account already known ultimate explanations of cooperation. And it is argued that Frank's model invokes a green-beard mechanism without acknowledging it (Gardner and West, 2009: p.33; West and Gardner, 2010, pp.1344, Heath and Rioux, 2018, pp.11-12). Hence, this model confronts the theoretical difficulty mentioned above. It is criticized for presenting a theoretically possible but not actually plausible ultimate explanation of human cooperation.

3. Robert Frank's Evolutionary Model and Social Behaviors

We have seen in the previous section that social evolution theorists rely on Hamilton's work to provide a coherent framework that categorizes social behaviors based on their consequences on direct fitness (Hamilton, 1964; 1970). As long as the phenomenon to be studied is clear, the condition that allows social behaviors to evolve is straightforward: an increase in the direct or indirect fitness of the social behavior relative to the average fitness in the population.

Inclusive fitness analysis can provide fresh insights into the ultimate explanations advanced by evolutionary models of social behavior. In particular, Gardner and West's analysis implies that the phenomenon explained by Frank's model is altruistic behavior. Moreover, this model applies a specific pleiotropy or linkage to explain the evolution of altruism. Hence, Frank's ultimate explanation appears to be in accord with key components of the green-beard explanation presented in the previous section.

The first task is to clarify the relevant social behavior explained in Frank's model. Since the stability concern applies to an ultimate explanation of altruistic behavior we cannot sweep conceptual issues under the rug and simply assume that insights from inclusive fitness analysis apply to Frank's model.

Robert Frank engages in evolutionary modeling to examine the evolution of social behaviors. He assumes that the context of interaction is predefined in such a way as

to pose a specific phenotype set (i.e. the strategies that could be applied in the game) and a payoff structure to a population of organisms. In particular, he considers a population whose members engage in a joint venture and the social interaction takes the form of a one-shot Prisoner's Dilemma. Instead of the usual terms "defect" and "cooperate", however, Frank's phenotype set consists of "dishonest" and "honest" behaviors. "To be honest here means to refrain from cheating one's partner in a cooperative venture, even when cheating be cannot punished. To be dishonest means always to cheat under the same circumstances" (Frank, 1987, pp.591). To examine the evolution of these behavioral types, Frank postulates the presence of phenotypic markers or signals indicating a partner's type. These markers are behavioral clues of emotions (e.g. respiration, posture, perspiration or facial muscle expressions) and follow a continuous probability density function.

This presentation might lead one to the conclusion that applying inclusive fitness analysis to examine Frank's evolutionary model is fundamentally flawed because there is a difference in the phenomena explained. One can emphasize that the Hamiltonian classification of social behaviors does not refer to honesty or trust which can mean many different things in the social sciences. Moreover, Frank's description of "honesty" (or related social behaviors like trust) does not refer to fitness effects and therefore these behavior cannot be interpreted as representing "altruism", "cooperation", or any related social behavior defined according to inclusive fitness analysis.

However, while plausible on the surface, these objections to the application of inclusive fitness analysis fail to be compelling. First, in Frank's verbal presentation of the prisoner's dilemma, cooperation, trust and honesty are often used interchangeably (e.g. Frank, 1987, 1994). One can further note that in the more detailed presentation of the model in the book "Passions within Reason", Frank states that the terms "defect" and "cooperate" are used to represent "dishonest" or "cheat" and "honest" or "not cheat" respectively (Frank, 1988, pp. 56). Second, it is possible to argue that Frank examines the same kind of behaviors as social evolution theory, even if he applies different labels. Although Frank's understanding of "honesty" is not based on fitness effects, the evolution of behavioral types in evolutionary game theory relies on material payoffs that represent or are related to fitness outcomes. Hence, one can assume that in his model Frank picks out a social trait that shares the same features as cooperation in social evolution theory. Honesty and cooperation refer to a individually deleterious behavior that benefits another organism.

So far I have left unexplored the problem of stability questioning evolutionary explanations that rely on the relation between the marker and social behavior. Let us first look deeper into the details of Frank's evolutionary model (Frank, 1987). In a footnote Frank writes that instead of honest (cooperative) and dishonest (defection) behaviors it is more accurate to consider two alleles a_1 and a_2 at a genetic locus that controls cooperation C and defection D respectively (ibid. fn. 7, pp.

595). In addition to differences in behavior, there are also differences with respect to the heritable component μ_i underlying some observable human feature S_i. One can further assume the presence of a second locus with two alleles, μ_H and μ_D. The different combinations of the four alleles imply that there are four distinct phenotypes $\{CS_H, DS_D, CS_D, DS_H\}$. In Frank's model, however, cooperation and defection are linked to different markers and the strategy set consists of $\{CS_H, DS_D\}$. It is easy to demonstrate that if cooperation C and defection D are associated with the same marker SH, and the strategy set consists of $\{CS_H, DS_H\}$, DS_H can invade the population and drive cooperation to extinction. Hence, Frank's modelling result depends on simply assuming away the presence of green-bearded defectors DS_H without providing any argument as to why such a constraint in the strategy set is plausible.

To sum up, Gardner and West's argument that Frank invokes a green-beard mechanism in the way defined by Hamilton and Dawkins does appear to follow from his evolutionary model. Although one can find different definitions of altruism, cooperation and honesty in Robert Frank's work, it is plausible to suggest that the behavior examined is actually altruism defined according to fitness costs and benefits. Moreover, what Gardner and West appear to analyze is Frank's phenotype set and his assumptions about the underlying genes. Cooperation and green-beard are correlated because of a linkage between the behavior and marker genetic loci. If the model allows for defectors to be associated with the same phenotypic marker that cooperators apply to discriminate against defectors (because of recombination or mutation), cooperation is not evolutionary stable.

In the section that follows I put aside Frank's evolutionary model and focus on the theoretical account presented in his book. At first sight, his evolutionary explanation on the evolution of emotion-based phenotypic markers appears to be very similar to the green-beard explanation of cooperation. Under close scrutiny important differences become apparent. My main aim in the section that follows is to point out what these differences are and examine their implication for the stability problem.

4. The Red-beard Evolutionary Explanation and Stability

Hamilton provided a general framework to study the evolution of cooperation, or any social behavior: Social behaviors are categorized based on direct fitness and are favoured by natural selection if they increase inclusive fitness. One can further distinguish between ultimate explanations of cooperation on the basis of direct and indirect fitness effects. In contrast to indirect ultimate explanations that rely on genealogical kinship, green-beard explanations are not evolutionary stable.

In this section I argue that it is possible to provide a satisfactory response to the stability concern, thus vindicating the basic thrust of Frank's account. However, in order to do so, one must examine in detail Frank's work on human emotions and

social behaviors. What I want to argue is that the standard green-beard explanation of social evolution theory and Frank's evolutionary model does not accord with many of the key components of the evolutionary explanation presented in the rest of Frank's work. While evolutionary models focus on fitness effects and expressed phenotype or strategies, evolutionary explanation can also refers to the emergence and evolution of psychological traits underlying social behaviors and phenotypic markers.

As a point of departure from the ultimate explanation provided by evolutionary models, one can consider that among the multiple definitions of social behaviors found in Frank's work, some are phrased in terms of proximate causes. Specifically, Frank suggests that cooperative individuals experience emotions that commit them to their threats and promises. He writes that "a cooperator is someone who, possibly through intensive cultural conditioning, has enhanced a genetically endowed capacity to experience a moral sentiment that predisposes him to cooperate. A defector is someone who either lacks this capacity or has failed to develop it" (Frank, 1988, pp.57).

While social evolution theory and inclusive fitness analysis focuses on behaviors and fitness outcomes and do not examine proximate causal processes underlying cooperation, Frank presents an unrefined description of cooperation in terms of emotions that are part of our psychological makeup. Cooperators and defectors are understood in terms of emotions which are proximate causes of social behavior. This understanding of cooperation is not uncommon in the social sciences (e.g. Gardner et al, 2007). While cooperation in evolutionary theory is characterized based on fitness outcomes, the social sciences often refer to the actor's internal processes that cause social behavior.

Moreover, it is possible to find a second aspect of emotions in Frank's account. Emotions such as sympathy not only motivate social behaviors like honesty, trust and cooperation in commitment problems such as the prisoner's dilemma, but also generate phenotypic markers that others can recognize. According to Frank, emotions manifest in physical markers such as facial expression, posture, voice, eye movements that convey a person's underlying emotion. These markers involve a suite of coordinated behavioral response patterns (i.e. skeletal, facial, vocal) that express internal emotion states. This link between human expressions and emotions facilitates communication between individuals, providing information about their internal states. Individuals often lack information about others' motives, which makes it difficult to determine their behavior and an appropriate course of action. This lack of insight of other individuals' desires, intentions or preferences was common in ancestral environments, where language as a communication device was not yet available.

What is Frank's rationale for describing human sociality in terms of psychological mechanisms? The reason, I argue, is that in the evolutionary explanation presented in Frank's book, proximate causes such as emotions are the target (or explanandum) of the evolutionary explanation. As argued by Frank, there is a strong relation

among emotions, expressions and social behaviors. In particular, social behaviors and phenotypic markers are tied to the same internal emotional states. And it is likely that humans have psychological capacities that evolved in the ancestral environment because the behaviors they sustained conferred material benefits related to fitness. According to Frank, "behavior is directly guided by a complex psychological reward mechanism" (Frank, 1988, pp.51) while "feelings and emotions, apparently, are the proximate causes of most behaviors" and the "task here, once again, is to explain how such sentiments might have evolved in the material world (ibid. pp.54).

Moreover, one can resist the conclusion that the phenomenon explained by Frank in his the book Passion within Reason is altruistic cooperation in the way defined by an inclusive fitness analysis. It is true that Frank states that like cooperation and honest behavior "will be one that, by definition, requires personal sacrifice" (Frank, 1988, pp.17). But he also adds that "if character traits like honesty are observable in a person, an honest person will benefit by being able to solve important commitment problems" (ibid. p.18). One can read in these passages that if a costly social behavior like honesty or cooperation is accompanied by a phenotypic marker, it ends up promoting an organism's individual fitness. In particular, costly social behavior and a marker end up conferring a direct fitness benefit on its bearers in comparison to self-interested behaviors with a different marker. And according to inclusive fitness analysis, such behaviors should be classified as mutually beneficial, not altruistic.

But why did emotions evolve? According to Frank, evolution would not have to build emotion-based mechanism from scratch in early human populations. Emotions related to social behaviors were initially favoured by natural selection not because of their physical manifestations, but because they contributed to the solution of problems related to human psychology. In particular, emotions may have first evolved as an impulse-control device guiding behavior towards long-term direct fitness benefits. They allowed individuals to undergo a cost in the present and gain higher fitness benefits in the future. Next, Frank suggests, observable markers followed the activation of these emotions. Once these markers emerged, they gradually became associated in the receivers' minds with the presence of emotions and natural selection refined them for communication purposes in one-shot interactions.

It is not my purpose to examine whether there is empirical evidence in support of Frank's account of the origin and evolution of emotions. What is important is that this evolutionary explanation differs from simple evolutionary model's ultimate explanation in the following way. In contrast to Frank's model and social evolution theory that focuses on behaviors and their fitness consequences, in the evolutionary scenario described in the previous paragraph the target of the explanation or explanandum is different. What is examined is the evolution by natural selection of emotions that produce social behaviors, a marker and the recognition of the marker. More generally, selection favoured the evolution of proximate mechanisms that

generate phenotypic markers and social behaviors. These proximate mechanisms have evolved because the behaviors they sustain provided, on average, fitness benefits. To distinguish this account from earlier discussions where the phenotypic marker is a green-beard, I call this the red-beard hypothesis.

Before we proceed to address the stability problem, two important clarifications are in order. First, the red-beard explanation does not a take a stance on the relation between our genetic architecture and proximate mechanisms. Despite scientific progress towards understanding human sociality, as things stand, we do not know the underlying genetics of human psychology and behavior to any level of precision (e.g. Reuter et al, 2010; Thompson et al., 2013). It is safe to assume that some genes feed into human proximate mechanisms and social behaviors. One can further add that our proximate mechanisms are likely to be polygenic. Or that an arrangement of some genes must have been present during the evolutionary processes that lead to the formation of the proximate mechanisms that sustain social behaviors and phenotypic markers. However, which particular alleles at specified loci regulate human proximate mechanisms and behaviors remains largely unknown. Given our ignorance, if we were to advance the hypothesis that proximate mechanisms are regulated by one gene or two loci and two alleles, this dependency on a specific arrangement of genes would undermine the credibility of the red-beard explanation right from the onset.

Second, the red-beard explanation directs our attention to the possibility that the actual total causal chain that leads to social behavior and the marker is different from the one presented by simple evolutionary models. It is possible to make this point clear if we consider what a red-beard explanation has to refer to. It will include (a) the gene(s) that shape emotion-based proximate mechanisms, (b) the two types of behavior produced by these mechanisms (in response to prevailing environmental influences), (c) the fitness outcome of these behaviors and (d) the influence of fitness on the dynamics of gene propagation. While a red-beard explanation focuses on the evolution of proximate mechanisms, simple evolutionary models like the one presented by Frank, black-box proximate causes and target patterns of behaviors and their relation to fitness. Moreover, it is implicitly assumed that the strong association between the marker and social behavior is either due to one gene feeding into the two behaviors or a linkage between the behavior and marker loci. In contrast, the red-beard explanation pays attention to evolution of emotion-based proximate causes in naturally occurring interactions which simple evolutionary models largely ignore. Social emotions like guilt, shame and empathy have a genetic basis, they involve hormonal and nervous system activities and these proximate causes connect social behaviors and markers that result to fitness effects.

What about the red-beard explanation? Could it be evolutionary stable? In the critique, Gardner and West are careful to avoid citing Passions within Reasons and focus on Frank's particular choices in model building (Gardner and West, 2009, pp.33; West and Gardner, 2010, pp.1344). In contrast to Gardner and West, a recent paper by Heath and Rioux does not distinguish between the explanation provided

by the Frank's evolutionary model and the one examined in his book (Heath and Rioux, 2018, pp.12-13)[31]. In highlighting the features of Frank's model, they observe that according to Frank, social emotions are posited as proximate causes that explain the relation between the marker and social behavior. They also add that it is more plausible to suggest that the falsebeard would arise in a population of cooperators with the marker, simply by losing or suppressing the capacity to cooperate (ibid. pp.11-12)[32] . What they fail to recognize is that these proximate causes are the target of the evolutionary explanation in Robert Frank's book and they ignore them when they examine stability issues.

In moving proximate causes such as social emotions to the center stage of analysis and viewing emotion-based mechanisms as factors that are significantly changed by genetic mutations allows us to address the problem of stability. What is crucial to observe is that the stability problem arises if cooperators and defectors end up having a very similar marker[33]. If, however, mutant genes cannot suppress cooperation without also suppressing the phenotypic marker, the stability problem does not arise.

Consider the following red-beard scenario. Suppose that genotype G feeds into an element of proximate mechanism E_M responsible for a social emotion E and this

---

[31]It is confusing that in the introduction of their paper, Heath and Rioux criticize the influence of the "first generation" socio-biology in evolutionary ethics, despite the fact that their arguments against Robert Frank's account are largely based on the inclusive fitness analysis found in the work of Gardner and West, two of the main exponents of Hamilton's revolutionary ideas in the field of sociobiology. More recent work in evolutionary social sciences examines the evolution of proximate mechanisms and attempts to understand human cognition and psychology by "looking into the black box" (e.g. Crawford & Krebs, 2008).

[32]Frank presents an alternative scenario, where defectors do not have capacity to produce the marker and the social behavior. Due to the complexity of the underlying proximate mechanism, Frank argues that it would be difficult for defectors to acquire the same marker that cooperators use to recognize each other (Frank, 1988, 2005). Heath and Rioux do not contest this argument: it is not very plausible that a random genetic mutation will result to the all-of-a-sudden emergence of the marker in those that do not have the underlying mechanism associated with a particular emotion.

[33]Cooperators and defectors do not need to end up having the same marker at all times. Provided, that the markers are different "on average", the proximate mechanism that produces the marker and cooperation could have evolved. It is also important to note that the production of emotional markers is not all or nothing. Their clarity or distinctness is a matter of degree. It is an empirical question whether there is the threshold between a particular marker being present or being absent.

emotion is a common cause of two correlated joint behaviors B and S. Although correlated, there is no direct causal relationship between B and S. As an illustration, E_M might refer to a proximate cause or mechanism related to an emotion E such as joy or sympathy, B is a social behavior like cooperation and S refers to a phenotypic marker such as a non-verbal expression of E. Cooperative behavior B is correlated with a marker S only because the proximate mechanism E_M is a common cause of both behaviors. This can be represented by a causal diagram in the following way:

$B \leftarrow E_M \rightarrow S$

Now assume that changing E_M does not have other effects to the organism besides changing B and S. Moreover, the pathway through which a mutant gene influences the social behavior B or the marker S goes through cognitive or psychological mechanisms E_M. Suppose that a genetic change happens due to mutation that suppresses social behavior B, as suggested by Heath and Rioux (2018). In this case, S will also change under this change of B. The reason is that changes to social behavior B are produced by changes to the proximate emotion-based mechanism E_M and this mechanism is causally related to both social behavior B and marker S. Hence, if one assumes that a mechanism E_M causes social behavior B and the marker S, whenever a change of E_M occurs due to mutation, it will result to changes both in the social behavior B and the marker S.

It follows that if emotions (or emotion-based proximate mechanisms) are taken into account and they causally produce social behavior and a phenotypic marker, cooperative behavior cannot be suppressed (or replaced by defection) without also suppressing (or losing) the phenotypic marker. In other words, those that possess or experience an emotion like sympathy that motivates cooperation will be "observably different, on the average, from those who do not" (Frank, 1988, pp.11). To the extent that psychological mechanisms of the kind that Frank proposes exist, Heath and Rioux's argument is unconvincing. Defectors will have a different marker than cooperators and cooperators will drive them to extinction.

One might think that the red-beard hypothesis underestimates the role of natural selection to suppressing the marker or/and the social behavior. In particular, the following two objections might come to mind. First, one might argue that natural selection would favour a mutation that suppresses both the marker and cooperation. Suppressing the whole emotion-behavior-marker complex would save the cost of providing benefits to other cooperators. However, this objection, fails to be compelling. In the red-beard explanation, bearers of a proximate mechanism related to an emotion are fitter than non-bearers (which is equivalent to defectors without a marker) and therefore individuals do not gain by suppressing both the social behavior and the marker. In other words, the presence of emotions attracts fitness benefits that outweigh fitness costs. If bearers of the emotion have higher fitness than non-bearers, a mutant gene suppressing the whole emotion-marker-social behavior complex will not be favoured by natural selection.

Secondly, one might object to the red-beard hypothesis by arguing that Frank's two-stage evolutionary account of emotions implies that natural selection resulted to a particular brain organization. Cooperators' brain end up having two more or less independent proximate mechanisms (or one proximate mechanism with two independent sub-mechanisms), each of which is devoted to the production of a particular kind of behavior (i.e. cooperation and the marker). This implies that natural selection could also act to change one mechanism independently of the other. In response to particular evolutionary pressures, natural selection would have favoured mutations that result in substantial changes in the proximate capacities underlying the marker production without influencing the proximate mechanism responsible for human cooperation.

While the analysis of emergence and modification of proximate mechanisms by natural selection is a complicated subject to which I cannot do justice here, such an argument is far more problematic than it first appears. The crux of this argument is the idea that two (or more) complex proximate mechanisms could have evolved and be selected for independently of one another. The problem with this argument is that it does not square well with what is known from evolutionary theory. Instead of furnishing largely independent proximate mechanisms, natural selection typically modifies the proximate mechanisms that were already present in the human brain[34]. Hence, one can retort that in the case of the red-beard hypothesis, a succession of genetic mutations over millions of years, whose phenotypes were subject to natural selection, resulted in one increasingly more complex proximate mechanism that ended up promoting non-verbal communication with others.

But even if one accepts that there are two proximate mechanisms (or a proximate mechanism with two sub-mechanisms) that underlie a particular emotion, it does not necessarily follow that natural selection can interfere with the operation of one proximate mechanism without influencing the operation of other. If the proximate mechanism involved in the production of the marker is built out of, and share parts with the proximate mechanism responsible for cooperative behavior, a red-beard explanation would involve partly distinct mechanisms. Natural selection might change those parts shared in common and a modification to the operation of one proximate mechanism might have an impact on the other[35].

---

[34] According to Ernst Mayr, "[T]he emergence of new structures is normally due to the acquisition of a new function by an existing structure" and "the resulting 'new' structure is merely a modification of a preceding" (Mayr, 1960, pp.377). More recently, evolutionary researchers such as Marcus (2006) and Anderson (2010) put forward similar ideas about proximate neural mechanisms, well supported by evolutionary arguments and empirical evidence.

[35] One way of thinking about this issue at the theoretical level is to apply the distinction between "mosaic" and "connected" traits (Sterelny and Griffiths, 1999). The red-beard hypothesis is based on the idea that many emotional traits are

The red-beard explanation can be readily understood by using animal signaling theory. Although the stability of signals can be maintained because of common interest, animal communication often involves a conflict of interest between signaler and receiver. In these cases, evolutionary theory suggests three main explanations why communication can be evolutionary stable (Maynard-Smith and Harper, 2003; Dezecache et al, 2013; McCullough and Reed, 2016): i) signals are indices, there are physiological, psychological, genetic or anatomical constraints that make it impossible to produce a deceptive signal; ii) signals are handicaps, namely there are differential costs associated with the production of the marker (Zahavi, 1975); iii) signals are stable because there are punishments costs incurred to dishonest signalers (Lachmann et al., 2001; Dezecache et al., 2013).

Given this distinction, in the red-beard explanation signals or markers are not handicaps. Nothing in the argument presented in the previous paragraphs requires that markers are costly to produce. What could be initially costly is social behavior related to the marker, although it ends up providing fitness benefits to the organism. Moreover, behavioral responses are unaffected by retaliation and the argument presented does not rely on potential punishment costs inflicted upon falsebeards.

In the red-beard explanation phenotypic markers are indices. According to the index explanation, stability is maintained because of the presence of a tight link between an underlying internal quality and its phenotypic expression. There is a causal relation between the marker and an underlying quality and the presence or intensity of the marker is related to that quality. However, the quality signalled is not a gene or a cluster of genes (the common cause in the green-beard hypothesis), but proximate behavior-generating emotions involved in social behaviors. What maintains cooperation is that the emergence of a defector with same marker as cooperators (i.e. “falsebeards”) is physically or psychologically difficult, because phenotypic markers are dependent on particular proximate cause of an organism.

To conclude, the red-beard and the green-beard can be held to differ with respect to what they explain. Frank’s evolutionary model and the green-beard explanation explain the evolution of social behavior (or phenotypic markers). The red-beard explanation, however, has a different target and explains the evolution of emotion, a proximate cause that commits individuals to social behaviors and produces a phenotypic marker. In both cases, what does the explaining is natural selection and the correlation between the marker and social behavior. The green beard explanation, however, is not stable because a mutant gene could arise that does not affect the marker but suppresses or replaces the social behavior. In contrast, a red-beard explanation is could be stable because a mutation will change both the marker and the social behavior.

---

connected traits. Hence, natural selection may not be able to modify the proximate mechanism responsible for a phenotypic marker independently of the (allegedly) distinct proximate mechanism that underlies human cooperation

5. Conclusion

This brief examination of Robert Frank's work has been, I think, adequate to establish the following: There are good reasons to focus on the evolution proximate causes since they often make an important difference in the stability or maintenance of social behaviors and phenotypic markers. The conventional view is that an acceptable ultimate explanation has to demonstrate that natural selection allows genes that encode for behaviors to spread, to sustain a sizable frequency in a population once they have spread and explicates why these behaviors would invade a population with alternative behaviors. While evolutionary models involve behavioral strategies and their fitness effects, we often need to examine the evolution of proximate mechanisms in complex evolutionary scenarios.

Targeting human behavior-generating mechanisms in evolutionary explanations opens the door to new research avenues that have been previously overlooked. First, the exact strategies and payoffs and the way they change depends on the internal condition of the player(s) involved. Because many of the results depend greatly on the strategy set, taking into account human psychology and cognition and expanding the phenotype set can also change the outcome of known evolutionary dynamics. Second, because proximate psychology may be at work within individuals during the evolutionary process and can influence and modify the behavior of conspecifics, populations starting with an interaction characterized by a game like the prisoner's dilemma can take an evolutionary path that leads to different game. Finally, an important direction for future research is to investigate the cognitive foundations of emotion recognition. Theoretical and empirical research could investigate whether distance in space has an influence on recognition costs. Even if in emotion-mediated cooperation individuals do not incur a cost for marker recognition, there is still the question why such costless human cognitive capacities emerged, evolved and stabilized by biological evolution.

References

Axelrod, R. (1984). The Evolution of Cooperation. New York: Basic Books.

Biernaskie, J. M., West, S. A., & Gardner, A. (2011). Are Greenbeards Intragenomic Outlaws? Evolution, 65(10), 2729–2742. https://doi.org/10.1111/j.1558-5646.2011.01355.x

Bshary, R., &Bergmuller, R. (2008). Distinguishing four fundamental approaches to the evolution of helping. Journal of Evolutionary Biology, 21(2), 405–420. https://doi.org/10.1111/j.1420-9101.2007.01482.x

Bylsma, L. M., Gračanin, A., & Vingerhoets, A. J. J. M. (2018). The neurobiology of human crying. Clinical Autonomic Research, 29(1), 63–73. https://doi.org/10.1007/s10286-018-0526-y

Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M., & Seabright, P. (2015). A Model of Smiling as a Costly Signal of Cooperation Opportunities. Adaptive Human Behavior and Physiology, 1(3), 325–340. https://doi.org/10.1007/s40750-015-0026-4

Cohen, E. (2012). The Evolution of Tag-Based Cooperation in Humans. Current Anthropology, 53(5), 588–616. https://doi.org/10.1086/667654

Crawford, C., & Krebs, D. (2008). Foundations of Evolutionary Psychology. Lawrence Erlbaum Associates/Taylor & Francis Group.

Darwin, C. (1872). The Expression of the Emotions in Man and Animals. London: John Murray.

Dawkins, R. (1976). The Selfish Gene. Oxford: Oxford University Press

Dawkins, R. (1982). The extended phenotype : the gene as the unit of selection. Freeman.

Dezecache, G., Mercier, H., & Scott-Phillips, T. C. (2013). An evolutionary approach to emotional communication. Journal of Pragmatics, 59, 221–233. https://doi.org/10.1016/j.pragma.2013.06.007

Dijk, C., Koenig, B., Ketelaar, T., & de Jong, P. J. (2011). Saved by the blush: Being trusted despite defecting. Emotion, 11(2), 313–319. https://doi.org/10.1037/a0022774

Ekman, P. (1993). Facial expression and emotion. American Psychologist, 48(4), 384–392. https://doi.org/10.1037/0003-066x.48.4.384

Frank, R. H. (1994). Group selection and “genuine” altruism. Behavioral and Brain Sciences, 17(4), 620–621. https://doi.org/10.1017/s0140525x00036244

Frank, R.H. (1987). If homo economicus could choose his own utility function, would he want one with a conscience? American Economic Review, 77. 593–604

Frank, R. H. (1988). Passions within reason : the strategic role of the emotions. Norton, Ca.

Frank, R. H. (2005). Altruists with Green Beards: Still Kicking? Analyse &Kritik, 27(1). https://doi.org/10.1515/auk-2005-0104

Gardner, A., & West, S. A. (2009). Greenbeards. Evolution, 64(1), 25–38. https://doi.org/10.1111/j.1558-5646.2009.00842.x

Grafen, A. (1990). Biological signals as handicaps. Journal of Theoretical Biology, 144(4), 517–546. https://doi.org/10.1016/s0022-5193(05)80088-8

Handfield, T., Thrasher, J., & García, J. (2018). Green beards and signaling: Why morality is not indispensable. Behavioral and Brain Sciences, 41. https://doi.org/10.1017/s0140525x18000080

Hasson, O. (2009). Emotional Tears as Biological Signals. Evolutionary Psychology, 7(3), 147470490900700. https://doi.org/10.1177/147470490900700302

Hamilton, W.D. (1964). The genetical evolution of social behavior. I & II. Journal of Theoretical Biology, 7, 1–52

Hamilton, W.D. (1970). Selfish and spiteful behavior in an evolutionary model. Nature 228, 1218–1220

Heath, J., &Rioux, C. (2018). Recent trends in evolutionary ethics: greenbeards! Biology & Philosophy, 33(1–2). https://doi.org/10.1007/s10539-018-9627-1

Izard, C. E. (2010). The Many Meanings/Aspects of Emotion: Definitions, Functions, Activation, and Regulation. Emotion Review, 2(4), 363–370. https://doi.org/10.1177/1754073910374661

James, S.M. (2011). An introduction to evolutionary ethics. Wiley-Blackwell, Malden

Joyce, R. (2006). The evolution of morality. MIT Press, Cambridge

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional Expression: Advances in Basic Emotion Theory. Journal of Nonverbal Behavior, 43(2), 133–160. https://doi.org/10.1007/s10919-019-00293-3

Lachmann, M., Számadó, S., & Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. Proceedings of the National Academy of Sciences, 98(23), 13189–13194. https://doi.org/10.1073/pnas.231216498

Marcus, G. (2006). Cognitive architecture and descent with modification. Cognition, 101(2), 443–465. https://doi.org/10.1016/j.cognition.2006.04.009

Maynard Smith, J., & Harper, D. (2011). Animal signals. Oxford University Press.

Mayr, E. (1960). The emergence of evolutionary novelties. In: Evolution after Darwin: the University of Chicago Centennial, vol. 1: The evolution of life, (ed.) S. Tax, pp. 349–80. University of Chicago Press.

Mayr, E. (1961). Cause and Effect in Biology: Kinds of causes, predictability, and teleology are viewed by a practicing biologist. Science, 134(3489), 1501–1506. https://doi.org/10.1126/science.134.3489.1501

McCullough, M. E., & Reed, L. I. (2016). What the face communicates: clearing the conceptual ground. Current Opinion in Psychology, 7, 110–114. https://doi.org/10.1016/j.copsyc.2015.08.023

Nesse R (2001) Natural selection and the capacity for subjective commitment. In: Nesse R (ed) Evolution and the capacity for commitment. Russell Sage Foundation, New York, pp 1–45

Owren M.J., Bachorowski, J.A. (2006). The evolution of emotional expression: a "selfish-gene" account of smiling and laughter in early hominids and humans. In: Mayne TJ, Bonanno GA (eds) Emotions: current issues and future directions. The Guilford Press, New York, pp 152–191

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. Trends in Cognitive Sciences, 17(8), 413–425. https://doi.org/10.1016/j.tics.2013.06.003

Reed, L. I., DeScioli, P., & Pinker, S. A. (2014). The Commitment Function of Angry Facial Expressions. Psychological Science, 25(8), 1511–1517. https://doi.org/10.1177/0956797614531027

Reed, L. I., Matari, Y., Wu, M., &Janaswamy, R. (2019). Emotional Tears: An Honest Signal of Trustworthiness Increasing Prosocial Behavior? Evolutionary Psychology, 17(3), 147470491987242. https://doi.org/10.1177/1474704919872421

Reed, L. I., Zeglen, K. N., & Schmidt, K. L. (2012). Facial expressions as honest signals of cooperative intent in a one-shot anonymous Prisoner's Dilemma game. Evolution and Human Behavior, 33(3), 200–209. https://doi.org/10.1016/j.evolhumbehav.2011.09.003

Reuter, M., Frenzel, C., Walter, N. T., Markett, S., & Montag, C. (2010). Investigating the genetic basis of altruism: the role of the COMT Val158Met polymorphism. Social Cognitive and Affective Neuroscience, 6(5), 662–668. https://doi.org/10.1093/scan/nsq083

Riolo, R. L., Cohen, M. D., & Axelrod, R. (2001). Evolution of cooperation without reciprocity. Nature, 414(6862), 441–443. https://doi.org/10.1038/35106555

Ruse, M. (1986). Evolutionary Ethics: A Phoenix Arisen. Zygon, 21(1), 95–112. https://doi.org/10.1111/j.1467-9744.1986.tb00736.x

Ruse, M. (1988). Evolutionary Ethics: Healthy Prospect or Last Infirmity? Canadian Journal of Philosophy Supplementary Volume, 14, 27–73. https://doi.org/10.1080/00455091.1988.10715944

Shariff, A. F., & Tracy, J. L. (2011). What Are Emotion Expressions For? Current Directions in Psychological Science, 20(6), 395–399. https://doi.org/10.1177/0963721411424739

Sterelny, K. (2016). Cooperation, Culture, and Conflict. The British Journal for the Philosophy of Science, 67(1), 31–58. https://doi.org/10.1093/bjps/axu024

Thompson, G. J., Hurd, P. L., &Crespi, B. J. (2013). Genes underlying altruism. Biology Letters, 9(6), 20130395–20130395. https://doi.org/10.1098/rsbl.2013.0395

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. The Quarterly Review of Biology, 46(1), 35–57. https://doi.org/10.1086/406755

Vromen, J. (2017). Ultimate and proximate explanations of strong reciprocity. History and Philosophy of the Life Sciences, 39(3). https://doi.org/10.1007/s40656-017-0151-4

West, S. A., & Gardner, A. (2010). Altruism, Spite, and Greenbeards. Science, 327(5971), 1341–1344. https://doi.org/10.1126/science.1178332

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. Journal of Evolutionary Biology, 20(2), 415–432. https://doi.org/10.1111/j.1420-9101.2006.01258.x

West, Stuart A., El Mouden, C., & Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. Evolution and Human Behavior, 32(4), 231–262. https://doi.org/10.1016/j.evolhumbehav.2010.08.001

Wilson, E. O. (1978). On human nature. Harvard University Press.

Zahavi, A. (1975). Mate selection—A selection for a handicap. Journal of Theoretical Biology, 53(1), 205–214. https://doi.org/10.1016/0022-5193(75)90111-3

Chapter 5: Hamilton's Rule: A Non-Causal Explanation?[36]

1. Introduction

Ever since Darwin (Darwin 1859, 1871), biologists have been interested in providing evolutionary explanations of social traits. Hamilton's (Hamilton 1963, 1964a, 1964b) groundbreaking work showed why costly social traits like altruism could evolve by natural selection. Although it was not clear what were the model's limitations and whether its assumptions were plausible, half a century of research has consolidated Hamilton's solution. Challenges from population genetics (e.g. Cavalli-Sforza and Feldman 1978; Karlin et al. 1983; Matessi and Karlin 1984) led to generalizations of inclusive fitness theory (IFT henceforth) (e.g. Queller 1992), to the examination of its relation to genetics (e.g. Karlin et al. 1983), and to evolutionary game theory (e.g. Taylor and Frank 1996).

Despite the immense theoretical progress and the development of sophisticated empirical methods during that same period, social evolution theory continues to be controversial. Notably, high-profile biologists Nowak, Tarnita, and Wilson (2010) joined forces in criticizing IFT, especially as it applies to microbes and social insects. Their *Nature* article provoked a fierce response of more than one hundred evolutionary biologists (Abbot et al. 2011) and subsequently generated a growing body of highly technical literature (e.g. Allen, Nowak, and Wilson 2013; Frank and Fox 2020; Gardner, West, and Wild 2011; Lehmann and Rousset 2014; Nowak et al. 2017; Rousset 2015; Veelen et al. 2017).

The exchange has also taken a philosophical twist beyond mathematical modelling or empirical research. Philosophers of science have attempted to clarify conceptual and methodological confusions that arise from different formulations of social evolution theory's main explanatory principle, Hamilton's rule (HR henceforth) (Birch 2014, 2017a, 2017b; Birch and Okasha 2015; Luque 2017; Okasha and Martens 2016; Rubin 2018). Within this literature, one key issue is whether HR *explains* and, if so, how. Our goal is to reinforce the case for the considerable value the general regression version of HR (Queller 1992; Gardner, West, and Wild 2011) has in explaining the evolution of social traits.

The chapter is structured as follows. In section 2, we introduce the controversy over HR's explanatoriness. Section 3 discusses one recent solution, namely Birch's (Birch 2017a, 2017b) proposal that HR explains qua being an organizing framework.

---

[36] Written together with Philippe Verreault-Julien. Both authors have contributed evenly to this chapter. A paper based on this chapter is currently under peer review.

Although we ultimately disagree with Birch that the rule explains for that reason, we agree that it does not explain by virtue of citing causes. In the following two sections, we argue that we can view HR as either providing an 'explanation by constraint' (Lange 2017, 2018b, see also 2013) or a non-causal counterfactual explanation (Jansson and Saatsi 2019; Reutlinger 2016, 2018; Saatsi 2018b; Woodward 2018). Section 4 examines the former and section 5 the latter. Section 6 concludes.

2. Does Hamilton's Rule Explain?

IFT is an extension of the genetic theory of natural selection. It is typically expressed by a mathematical inequality called HR. The rule states the conditions under which genes for various types of social traits (i.e. social phenotypes) will spread by the evolutionary process of natural selection. HR demonstrates that whether a social trait undergoes selection hinges on the genetic relatedness between those that interact and by the changes in their expected number of offspring.

Depending on the evolutionary problem at hand, different forms of HR can be derived from evolutionary game theory or principles of population genetics. From a philosophical perspective, this is crucial for understanding the current debate. When evolutionary biologists discuss HR, they may have quite different principles in mind. The original formulation of HR states that an altruistic trait will see its frequency increase if and only if relatedness r times the trait's benefits b is greater than the trait's cost c, or $rb > c$ (Hamilton 1964a, 1964b). Hamilton derived his result based on several assumptions such as one-locus population, no epistasis (or dominance), weak selection, and fitness additivity. Following that result, Hamilton and numerous other evolutionary theorists relaxed the original assumptions and examined whether they could recover a similar result with, e.g., class structure, kin competition, or non-additive fitness payoffs (e.g. Queller 1985, 1992; Hamilton 1970). Hence, more than five decades of extensive theoretical research resulted in a variety of formulations of the type $rb > c$ (e.g. Frank 1998, 2013), often coming with different assumptions under which HR applies (e.g. Frank 1998; Lehmann and Rousset 2014; Queller 1984, 1992, 2011).

According to Birch and Okasha (2015; see also Birch 2014), depending on the meaning of 'cost' and 'benefit,' it is possible to distinguish among three main versions of HR: general, special, and approximate. A general version (HRG) is derived from the Price equation (Queller 1992), with the costs and benefits corresponding to partial regression coefficients. A special version (HRS) emanates from evolutionary game theory (Queller 1984; Veelen et al. 2012), with costs and benefits reflecting the payoffs of a game matrix. In the approximate formulation (HRA), the costs and benefits of the regression coefficients are approximated by partial derivatives (Taylor and Frank 1996).

Against this background, both the philosophical and the biological literature have mainly focused on the general version of the rule derived from the Price equation (Price 1970, 1972). As a mathematical model, the Price equation can be applied to different kinds of evolutionary processes, both biological and non-biological. In biological applications, it is used to provide a representation of the process of natural selection by the covariance between fitness and the trait of interest (e.g. Frank 2012; see Luque 2017 for a review of extensions and applications).

According to the Price equation, for a specific trait$g$, the average change in the value[37] of the trait $\Delta\bar{g}$ is:

$$\Delta\bar{g} = \frac{1}{\bar{w}}[\mathrm{cov}(w_i,g) + E(w_i\Delta g_i)]$$

Derivations of HRG often assume that mutation bias, genetic drift, and meiotic drive do not influence a trait's spread and that, therefore, the second term of the Price equation is zero. What remains captures the action of natural selection, i.e. the change in the genetic value of the social trait.

$$\Delta\bar{g} = \frac{1}{\bar{w}}[\mathrm{cov}(w_i,g)]$$

A general version of HR can be derived by expressing fitness and traits as linear regressions on an underlying heritable genetic element and by combining this regression with the Price equation (Gardner, West, and Wild 2011; Frank 1998; Queller 1992). A linear regression equation expresses fitness in terms of statistical coefficients. In the case of neighbour-modulated fitness,[38] we have:

$$w = \bar{w} + \beta_{wg.g'}(g - \bar{g}) + \beta_{wg'.g}(g' - \bar{g}') + \varepsilon$$

Here $\beta_{wg.g'}$is the partial regression coefficient of fitness w on one's genes g, holding the neighbor's genes g' constant, while $\beta_{wg'.g}$ is the partial regression coefficient of w on g', holding g constant. Substituting the regression equation (2) into the Price equation and taking into consideration that in least squares analysis $\mathrm{cov}(g,\varepsilon) = 0$ results in an equation that describes average genetic change in terms of partial regression coefficients.

---

[37] Typically it refers to the change in the individual's allele frequency at a locus or a linear recombination of these frequencies (i.e. breeding value).

[38] A way of partitioning fitness into a direct fitness component attributed to behaviour of the focal individual and an indirect fitness component ascribed to the behaviour of its neighbours.

$$\Delta\bar{g} = \frac{1}{\bar{w}}[\beta_{wg.g'}(\mathrm{cov}(g,g) + \beta_{wg'.g}\mathrm{cov}(g,g')]$$

Since $\mathrm{cov}(g,g')/\mathrm{cov}(g,g)$ is equal to$\beta_{g'.g}$, for $\mathrm{var}(g) \neq 0$ we obtain HRG.

$$\Delta\bar{g} > 0 \quad \text{iff } \beta_{wg.g'}\beta_{g'.g} + \beta_{wg'.g} > 0 \text{ or } rb - c > 0$$

In this general form of HR, costs and benefits are conceived as least-squares partial regression coefficients of the individual's fitness on its own and its partner's genetic values (e.g. Queller 1992; Gardner, West, and Wild 2011). In this formulation, $r = \mathrm{cov}(g,g')/var(g)$ is the regression coefficient of relatedness, a statistical concept typically understood as the extent to which individuals carry the same genes relative to all individuals in the population, $-c$ is the fitness cost of carrying the gene for a social trait, and b is the fitness benefit provided to social partners. In other words, HRG mathematically separates direct fitness $-c$ from the indirect fitness rb.

A point of contention concerns the explanatory power of HRG (e.g. Allen, Nowak, and Wilson 2013; Nowak et al. 2017; Veelen et al. 2017). Nowak et al. (2017, 5669), for example, question the validity of HRG and claim that "in its exact and general formulation […] [HRG] neither predicts nor explains the evolution of social behavior." The issue here is that partial regression coefficients and, therefore, the costs and benefits terms of HRG, cannot be interpreted as causes since, as is well known, correlation does not imply causation. And if costs and benefits lack causal meaning, how could HRG explain?

There have been two main attempts to rescue HRG's explanatoriness, namely that:

1. It provides a causal explanation if the regression coefficients have causal meaning (Okasha and Martens 2016).

2. It constitutes a 'unifying principle' (Birch 2014) or 'organizing framework' (Birch 2017b).

The first strategy is straightforward. The causal account of explanation is widely considered to be successful (e.g. Strevens 2008; Woodward 2003). Hence, if it can be shown that HRG cites the causes of phenomena, then its explanatoriness is easily accounted for; it is just a regular causal explanation. The problem, as we will discuss in more detail in section 5, is that regression coefficients are not always causally interpretable. In a nutshell, Okasha and Martens (2016) show that in the case of non-additive pairwise interactions with synergistic effects, attributing causal meaning to HRG requires particular — and controversial — assumptions about environmental constancy.

The second strategy is more subtle and involves showing that HRG explains, but not because it cites the causes of social evolution. Birch (Birch 2017a, 2017b) argues that HRG explains because it is an *organizing framework*. HRG does not have the same explanatory role as more detailed evolutionary models. Instead, it facilitates

evolutionary research by organizing causal explanations into meaningful classes. According to Birch, HRG allows the interpretation, classification, and comparison of evolutionary models that make substantial assumptions about social evolution.

As Birch (2017b, 50) observes, this is not a totally novel suggestion. It is common for evolutionary theorists to classify and interpret the results of evolutionary models within the framework provided by IFT (e.g. West, El Mouden, and Gardner 2011). Evolutionary theorists often aim for generality instead of application to particular problems (e.g. Queller 1992; Gardner, West, and Wild 2011). A general formulation of HR that does not depend on restrictive assumptions is the appropriate tool for conceptual analysis. According to Gardner and his colleagues, HRG can be applied…

> […] as a conceptual aid for understanding the results delivered by these analyses [mechanistic, dynamically sufficient models], and connecting these with the results of other models analysed using different methods, as it provides a general framework and universal language for social-evolution theory (Taylor and Frank 1996; Frank 1998; Rousset 2004; Gardner, West, and Barton 2007). (Gardner, West, and Wild 2011, 1037)

Although we think these two strategies may be correct, we also believe they do not tell the whole story for two reasons. First, as Birch and Okasha (2015) acknowledge, whether HRG can receive a causal interpretation is a point of contention. In particular, a causal interpretation of HRG is uncontroversial only in the case of non-additive effects (Okasha and Martens 2016; Veelen et al. 2017). That puts into question the merits of pursuing that strategy. Second, even if grant that HRG is an organizing framework, we do not believe that it explains for *that* reason. This is the subject of the next section.

3. The organizing framework defence

Birch (Birch 2017a, 2017b, see also 2014) holds that HRG is an organizing framework that provides a 'classificatory scheme' that allows to interpret all models using the $rb > c$ inequality. As such, HRG may classify as similar particular models that cite categorically different mechanisms. For instance, mechanisms of kin discrimination (via environmental or genetic cues) and limited dispersal (population viscosity) have in common that they promote the evolution of social traits by generating a sufficiently high degree of relatedness between interacting organisms. However, this same result is achieved through distinct causal routes. According to Birch (2017b), other models instead emphasize that some traits have a negative cost ($-c$), thus downplaying the need for a positive rb; social traits may evolve due to their direct fitness effects. Similarly, various causal routes may lead to negative costs. HRG, Birch says, thus allows to classify the models used to represent all evolutionary target systems.

He gives the following threefold account of an organizing framework. Suppose there is a set of evolutionary models M of ecological scenarios S. A model Ω, not itself in M, is an organizing framework if:

1. Ω represents all the target systems represented by the members of M, but does so in less detail,

2. Ω assumes nothing that is not assumed by all members of M, and

3. the relations between Ω and the members of M enable researchers to classify the members of M in an illuminating way. (Birch 2017b, 48–49)

Birch argues that HRG satisfies the first condition because we can view it as an abstract representation of evolutionary change for any genetically inherited trait. It meets the second one because HRG shares some key assumptions with a wide range of evolutionary models that represent change due to natural selection acting in a constant genic environment rather than the change attributed to genetic drift or meiotic drive. Finally, it fulfils the third condition because HRG's partitioning of evolutionary change into rb and $-c$ components allows to classify explanatory models by how these components interact to bring about evolutionary change. Hence, Birch holds that HRG has the three key features of an organizing framework.

It is true that in its general formulation, HR transcends particular evolutionary models that analyze the genetic evolution of social traits and facilitates contemporary mathematical research on social evolution. If a trait is analyzed correctly, researchers should expect that it would evolve under the conditions specified by HRG. To give an example, over the last decades researchers have advanced a series of strong reciprocity models that explain social behavior in the laboratory or the field (e.g. Bowles and Gintis 2004; Gintis 2000). The claim is that these models provide a distinct and novel account for the evolution of cooperation that is outside the confines of evolutionary explanations provided by IFT. Strong reciprocity refers to a combination of norm-abiding cooperative behaviors and costly punishment of norm violations. However, by distinguishing punishing from cooperative behaviors, West, El Mouden, and Gardner (2011) and Vromen (2012) demonstrate that strong reciprocity could not have evolved unless it provided either direct or indirect fitness benefits. In many cases, punishment behavior may provide a direct fitness benefit by increasing cooperation within a group. It can be also the case that it may result in indirect fitness returns by decreasing the fitness of competing non-genetic relatives. Hence, this modelling literature does not identify a novel pathway to the evolution of social traits. Instead, it provides an indirect fitness explanation, a direct fitness explanation, or a combination of both (Birch 2017b).

Before going further, we should say we agree with Birch on two points. First, that HRG can serve as an organizing framework. Birch's case is convincing and consistent

with our reading of the practitioners.[39] Second, we also believe that HRG explains and, in particular, that it provides a non-causal explanation.[40] However, pace Birch, we do not think it explains qua organizing framework. Importantly, these two views can coexist: HRG can be an organizing framework and explain for a *different reason* than being an organizing framework. In other words, our contention is that HRG explains on grounds other than those identified by Birch. Before looking at them, let us examine his claim.

He provides two main arguments for HRG's explanatoriness. The first consists in inferring that it explains on the basis that it affords understanding. He says that HRG "would be non-explanatory only if it added nothing of value to our understanding of social evolution" (Birch 2017b, 69). According to him, HRG is 'unifying' in a way that existing philosophical theories of explanation do not capture (e.g. Birch 2017b, 68, fn. 4). In particular, Birch notes that HRG does not unify in Kitcher's (Kitcher 1981, 1989) sense of providing argument patterns.

Even though we are sympathetic to this line of argument, whether having understanding implies having an explanation (i.e., explanation is necessary for understanding) is a matter of dispute (e.g. Dellsén 2020; Gijsbers 2013; Lipton 2009; Rice 2016). There are two reasons to question that connection between understanding and explanation. First, it is plausible that we can obtain explanatory understanding *without* having an (actual) explanation. Some have argued that so-called how-possibly explanations, viz. explanations that do not actually explain phenomena, may afford understanding (e.g. Rohwer and Rice 2013; Ylikoski and Aydinonat 2014).[41] While (actual) explanations may be an important source of understanding, that literature suggests it is not the only route. Likewise, Lipton submits that "we may unify the phenomena (and so improve our understanding of them) by constructing schemes of classification that do not in themselves provide explanations […]" (2009, 54). According to him, Kuhnian exemplars are one plausible mechanism through which we can acquire that knowledge of unification without explanation.

---

[39] Levin and Grafen (2019) explicitly endorse the organizing framework interpretation, but they do not discuss it in the context of explanatory considerations.

[40] Birch (2017b, 76) also suggests that under a liberal interpretation of 'causal explanation' we could say that HRG explains causally. We prefer to stick with the non-causal label since the liberal interpretation involves a significant departure from standard accounts. See also Andersen (2018) for a discussion of broad and narrow construals of causal explanations.

[41] This also suggests we could interpret the disagreement between the critics and supporters of HRG as one concerning the explanatory status of HRG, i.e. whether it provides an *actual* or a *possible* explanation.

Second, it is not clear whether the sort of understanding HRG affords qua organizing framework is properly explanatory. Gijsbers (Gijsbers 2013, see also 2014) proposes a distinction between two sorts of understanding, viz. explanation-understanding and unification-understanding. According to Gijsbers, only the former is explanatory in virtue of identifying what he calls "vertical connections" between phenomena. By contrast, one gains unification-understanding by knowing about the "horizontal connections of kinship" between phenomena. Knowing horizontal connections allows one to see the significant similarities between phenomena; it unifies them. Gijsbers claims that this form of unification affords understanding, yet does not explain. More generally, the worry is whether or not all understanding is of the explanatory sort. In particular, some have argued for an objectual (vs explanatory) notion of understanding (e.g. Dellsén 2020; C. Z. Elgin 2017; Kelp 2015). Therefore, if we cannot dismiss the possibility that we gain understanding without explanation and if understanding is not necessarily of the explanatory sort, then we cannot straightforwardly infer from the benefit of understanding to the conclusion that HRG actually explains. These debates indicate that whether HRG affords understanding is not, by itself, sufficient to warrant the conclusion that it explains. As a result, we need to look in more detail at the other reasons Birch provides.

This is what the second argument for HRG's explanatoriness does. Let us then assume that it is correct to infer explanatoriness from the presence of understanding. Why does HRG afford understanding in the first place? Of the three conditions Birch states for a model Ω to serve as an organizing framework, we take it that the third one is the most important with respect to its explanatoriness. This is because it would be possible to build an arbitrary framework Ω that satisfies conditions (i) and (ii), but without any explanatory value. What matters is that Ω's classification of models M explains. For Birch, HRG explains because it "generates understanding of causes by providing a framework for comparing, classifying, and interpreting such explanations" (Birch 2017b, 76). Elsewhere, he also says that "[t]he insight embodied by HRG is that every adequate causal explanation of positive change can be placed somewhere in this space" (Birch 2017a, 4). In short, the point is that HRG provides an illuminating way to classify evolutionary explanations and since that classification affords understanding, then it also explains. For instance, HRG may help to see that different causal mechanisms increase rb and therefore have the same effects. Or, if we observe that a given trait is costly, then we can infer that it must be explained by its indirect fitness effects rb.

We agree that HRG provides an illuminating classification of evolutionary models. However, explanations are typically judged by their capacity to increase our understanding of empirical phenomena of interest. According to Birch, what HRG helps us to understand are the various evolutionary explanations it organizes. It provides insights on evolutionary theory by organizing explanations on the basis of their commitment to the sign of r, b, and c (Birch 2017b, 46ff.). Indirect fitness explanations, for instance, "rely on there being a mechanism that explains the systematic tendency for the benefits caused by the expression of genes for altruism

to fall differentially on other bearers of those genes […]" (Birch 2017b, 51). Limited dispersal can be such a mechanism that explains why the benefits caused by altruistic genes are differentially directed to those that have those genes. HRG shows that all these indirect fitness explanations share in common $rb > 0$ and $c \geq 0$.

The main issue with this argument is that what we appear to understand better with HRG is not empirical *phenomena*, but social evolution *theory*. Put differently, HRG qua organizing framework affords theoretical, not explanatory understanding.[42] De Regt's (De Regt 2017, see also 2009; De Regt and Dieks 2005) distinction between understanding a theory and understanding phenomena is helpful to make this point plain. According to him, a "phenomenon P is understood scientifically if and only if there is an explanation of P that is based on an intelligible theory T and conforms to the basic epistemic values of empirical adequacy and internal consistency" (De Regt 2017, 92). Intelligibility is a pragmatic notion and refers to a cluster of virtues that theories possess. It is thus context-dependent in that different epistemic communities may assess the intelligibility of given theories differently.

What is important for our purposes is that although having an intelligible theory is a necessary condition for understanding phenomena, it is not sufficient. Scientists also need to use the theory to construct (correct) explanations. We take it that what HRG does qua organizing framework is to improve the intelligibility of social evolution theory. As such, it facilitates the construction of explanations. But theories and explanations are not the same things and facilitating the construction of explanations is not the same as explaining. The explanations explain, not the intelligible theory. In De Regt's terms, while understanding a theory is a condition for understanding phenomena, understanding phenomena is the "product of explanations" (2017, 92). Explanations, in turn, may have to satisfy different conditions. For instance, proponents of causal explanations hold that all explanations must cite causes. Others are pluralist. De Regt himself endorses a rather minimal set of conditions by saying that adequate explanations only need to be empirically adequate and internally consistent. If this is correct, then what would explain are the explanations we can construct using HRG, e.g. indirect fitness explanations. So HRG could be instrumental in building explanations, but in and of itself would not explain.

At this point, one could retort that if an organizing framework improves the intelligibility of social evolution theory, then does it not also help to understand phenomena? For if understanding a theory is necessary for understanding phenomena, then surely improving our grasp of social evolution theory is instrumental to our understanding of social evolution phenomena. We have nothing against this conclusion. It is plausible to suggest that the degree of understanding

---

[42] Yet other ways to frame the problem would be to say that HRG affords understanding-with (Strevens 2013). See Newman (2017) for an account of theoretical understanding.

phenomena depends on a theory's intelligibility (see De Regt 2017, 100, fn. 10). However, our main contention stands, namely that despite HRG's positive contribution to our understanding of social evolution *theory*, that does not make it an *explanation* of social evolution *phenomena*. We thus believe that if HRG explains, it must be for reasons other than being an organizing framework.

4. An explanation by constraint

The dominant account of explanation is the causal one (Reutlinger 2017; see also Strevens 2008; Woodward 2003). It is thus not surprising that evolutionary researchers have aimed at showing how HRG may identify the causes of social behavior (e.g. Okasha and Martens 2016). However, a growing literature has challenged that predominance by investigating putative cases of non-causal explanations, in particular mathematical explanations (Baron 2019; Baron, Colyvan, and Ripley 2017; Jansson and Saatsi 2019; Lange 2013; Pincock 2015; Räz 2017; Reutlinger 2016; Saatsi 2012).[43] Basically, what these authors argue is that there are bona fide explanations that explain despite them not citing causes.

In what follows, we will explore two types of accounts that illuminate the debate over HRG's explanatoriness. The first type are 'explanations by constraint' (Lange 2013, 2017, 2018b). This will be the object of this section. The second type is what we call the non-causal counterfactual account (e.g. Jansson and Saatsi 2019; Reutlinger 2016, 2018; Woodward 2018), which we discuss in section 5. We use both of these accounts for two main reasons. First, we do not want to take a stand concerning whether there is a single best account of non-causal explanations or whether one is superior to the other. Second, if we can show that there are plausible readings of HRG as either type of non-causal explanation, then this strengthens our claim that we can, indeed, interpret HRG as providing a non-causal explanation.[44]

Explanations by constraint show that the explanandum is due to facts that are more necessary than ordinary causal generalizations. Consider the following simple example.

---

[43] Here, 'mathematical explanations' means scientific explanations of empirical phenomena that explain in virtue of — non-causal — mathematical facts. We are not here referring to explanations *within* mathematics (see Mancosu 2001; Steiner 1978).

[44] Whether explanations by constraint and non-causal counterfactual explanations are different species is contentious. For instance, Saatsi (2018a) considers that explanations by constraint can be subsumed under the more general counterfactual umbrella. Our argument does not hinge on that debate since our goal is to show that HRG explains according to either one of these accounts.

> The fact that 23 cannot be divided evenly by 3 explains why Mother fails every time she tries to distribute exactly 23 strawberries evenly among her 3 children without cutting any (strawberries—or children!). (Lange 2017, 6)

Lange argues that what explains Mother's failure to divide evenly the strawberries is not the particular physical facts of the situation. Rather, it is the modally stronger mathematical facts. It is just impossible to divide evenly 23 (whole) units of anything. Although we could make the division possible by altering the physical conditions (e.g., by distributing 24 strawberries), the facts that prevent (or would allow) the division are mathematical, not physical.

Lange holds that many scientific explanations function similarly by appealing to 'constraints.' He says that these explanations…

> […] work not by describing the world's network of causal relations in particular, but rather by describing the framework that any physical system (whether or not it figures in causal relations) must inhabit, where this variety of necessity is stronger than the necessity possessed by the ordinary laws of nature. (Lange 2017, 44; see also 2013)

For Lange, explanations by constraint are non-causal since what explains are not causal facts, but facts — constraints — of a stronger modality. A crucial aspect of Lange's account is the idea that there are different levels of necessity (see also Lange 2009). While ordinary causal generalizations could easily have been otherwise, some laws or principles are modally stronger. These can be logical or mathematical facts, but also contingent laws of nature like, for instance, Newton's second law of motion. What matters is that their modal force transcends that of weaker causal facts. Constraints, then, just are facts that manifest a stronger necessity than the particular causal details of a situation.

For example, in the case of Mother and her strawberries, we could explain her failure by citing her repeated and unsuccessful attempts at sharing evenly the 23 strawberries with her 3 children. The causal process underlying the distribution of the strawberries does not bring about even shares between the children. But this causal process is modally weaker than the mathematical facts. Thus, these facts constrain the causal details. Regardless of how Mother would try to distribute the strawberries, she would fail. Hence, if the explanandum occurred it is *because of* these constraints, not because of the causal details.

We believe that HRG can be interpreted as explaining in virtue of being a constraint. HRG transcends causal explanations because it delineates the biological (causal) space of possibility. It provides a formal and conceptual framework that all causal explanations need to satisfy. Let us illustrate this with the following why-question:

Why has altruism never evolved without positive relatedness?[45] According to the standard definition of evolutionary (or biological) altruism, it is a behavior that results in a decrease in the actor's lifetime fitness and in an increase of the recipient's lifetime fitness. We could explain this fact by providing the causal details of the different situations, for instance by showing that kin discrimination was absent and thus could not result in positive assortment. Or, we could show that in another case the lack of positive assortment was due to an absence of phenotypic markers. By doing this we would answer the why-question by showing how the organisms were not positively related in virtue of the causal details. In other words, we would provide a causal explanation. But we can also explain why altruism has never evolved without positive relatedness by appealing to HRG. As we have seen in section 2, if $\Delta\bar{g}$ denotes the average change in the frequency of a gene between the ancestral and descendant populations, natural selection favours genetically transmitted social traits when:

$$\Delta\bar{g} > 0 \quad \text{iff} \quad rb - c > 0 \text{ given that } var(g) \neq 0$$

What this implies is that regardless of the social trait under examination, how the population is structured, or how the trait results to costs and benefits, it was selected $because rb > c$. Benefits b, costs c, and genetic relatedness r are all regression coefficients and thus might not represent causal processes (see section 5 below). Although r can be negative in the case of spiteful behavior, altruism cannot evolve if relatedness is equal to zero. For if r is zero, then the indirect fitness term rb equals zero too. This leads to the mathematically impossible result $-c = \beta_{wg.g'} > 0$ (a negative number cannot be higher than zero). At a formal level, HRG requires a positive r for altruism to evolve.

Therefore, altruism could not have evolved in the absence of positive relatedness simply because it did not fulfil the conditions stated by HRG. Crucially, HRG holds no matter the causal details. For instance, if altruism evolved in a population, HRG would have still held if different causal processes had been responsible for the evolution of altruism. As long as behavior comes at a net cost for the actor and net benefit for the recipient, relatedness needs to be positive. The particular details giving costs, benefits, and relatedness their values are irrelevant. In that sense, HRG acts as a constraint since it is modally stronger than the causal generalizations it applies to; it constrains potential causal explanations of the trait's evolutionary

[45] It would also be possible to formulate the why-question in a slightly different way: Why is it impossible for altruism to evolve without positive relatedness? We chose the former formulation because the explanandum is clearly empirical whereas in the latter it would be a modal fact, which might — incorrectly — suggest that HRG does not explain biological phenomena. That said, either formulation is amenable to an explanation by constraint, so our argument does not hinge on that (Lange 2018b).

change. *Any* causal explanation of social traits will be consistent with the $rb - c > 0$ inequality. What practitioners say supports this interpretation.

> It is simply incorrect to claim that Hamilton's rule requires restrictive assumptions or that it almost never holds (e.g. Veelen 2009; Nowak, Tarnita, and Wilson 2010). […] [HRG] *applies to any scatter of genetic and fitness data, irrespective of the underlying causes of this variation*. It makes clear that, whatever the relationship between trait and reproductive success, the resulting action of natural selection can be decomposed into cost, benefit and relatedness terms. (Gardner, West, and Wild 2011, 1037, our emphasis)

Gardner, West, and Wild make two claims of particular interest. Firstly, they reject the interpretation according to which HRG rarely holds (cf. Nowak, Tarnita, and Wilson 2010). In other words, HRG actually explains and applies to any data. Secondly, they assert that HRG holds *regardless* of the particular causal details. We take this particular point to be crucial for our claim that HRG explains qua constraint. HRG's explanatory contribution does not lie in its specification of the particular causes of the phenomena it purports to explain. Instead, it applies to all fitness and genetic data, irrespective of the causal details. No matter how one wants to explain the evolution of particular social traits, the relationship between benefits, relatedness, and costs will be related as per HRG states. In fact, not only HRG does not specify how individuals are causally related in a given population, but for HRG these causal details are irrelevant. In a given explanatory context, HRG implies there is a degree of necessity that goes beyond what ordinary causal explanations would tell us.

One could now ask whether explaining qua constraint is really different than explaining qua organizing framework since an organizing framework arguably sets boundaries to causal explanations. Our reply is twofold. First, the explanation by constraint account emphasizes the actual feature that is responsible for HRG's explanatoriness. Being an organizing framework is accidental to whether a principle like HRG explains. Some organizing frameworks explain because they constrain; some constraints are not organizing frameworks. We identify what is actually responsible for HRG's explanatoriness, namely that it acts as a constraint. Organizing frameworks and constraints are not coextensive. For example, a classification of mammals arguably satisfies all of Birch's (2017b) criteria for an organizing framework, but does not explain. Conversely, some laws of nature (e.g. Newton's second law, see Lange 2017) may act as constraints without being an organizing framework in the sense of Birch (2017b). Second, as we have said above, it is unclear how HRG qua organizing framework explains phenomena since its main function is to classify *explanations*. Whereas HRG qua constraint tells us why altruism evolved (because of positive relatedness), an organizing framework does not provide such answers. Therefore, we believe that the explanation by constraint account better locates the source of HRG's explanatoriness.

It may also be possible to object that there is an important disanalogy between HRG and the case of Mother and her strawberries. Whereas the latter invokes mathematical facts, HRG describes the conditions under which natural selection will favour social traits. It is relatively uncontroversial that mathematical facts are modally stronger than causal ones, but HRG does not seem to fit squarely in either category.[46] Although delving into the details of Lange's (see Lange 2009, 2017) account of degrees of necessity could be interesting, the only important thing for our purposes is that HRG is modally stronger. Current interpretations from practitioners and philosophers alike support this weak claim. Indeed, they often express HRG's contribution with a biconditional relation between the explanandum and explanans. Birch's (Birch 2017a, 2017b) characterization is explicit: $\Delta\bar{g} > 0$ iff $rb - c > 0$. Likewise for Gardner and West (2014; 1, our emphasis; see also Gardner 2015) who write that "[HRG] states that any trait—altruistic or otherwise—will be favoured by natural selection *if and only if* the sum of its direct and indirect fitness effects exceeds zero." Causal explanations typically do not identify necessary conditions. There is a multitude of causes that may be responsible for the evolution of altruistic traits: limited dispersal (i.e population viscosity) keeps relatives together or kin discrimination based on environmental or genetic cues typically generates a high degree of genetic relatedness among interacting individuals. Both, therefore, favour the evolution of altruistic cooperation. What causes the costs, benefits, and relatedness to have particular values may be different across populations. The key point is that regardless of the causes at play, the relationship between r, b, and c will hold. In that sense, it is more necessary than the particular causes.

Furthermore, both champions (e.g. Frank 2012) and critics (Allen, Nowak, and Wilson 2013; Nowak et al. 2017; Veelen et al. 2012) alike point out that HRG is akin to a tautology. If HRG is indeed modally stronger than the causal details it applies to, then we should expect HRG to appear as such. Moreover, we should also expect that to refute it would require meeting a different set of conditions. Whereas we can see the inadequacy of a causal explanation by the lack of a causal relationship between the explanans and the explanandum, this would not falsify the constraint. The constraint holds regardless of the causal details. This is why we take HRG's apparent tautological nature as evidence for our claim that HRG has a stronger degree of necessity than particular causal explanations of the same explanandum. HRG thus lays out a mathematical inequality that all causal explanations need to satisfy. Regardless of how one wants to causally explain particular evolutionary phenomena, the relationship between benefits, relatedness, and costs will abide by HRG.

---

[46] Here we sidestep the debate over whether the principle of natural selection is a priori or not (M. Elgin and Sober 2015; Lange and Rosenberg 2011; Sober 2011).

5. A non-causal counterfactual explanation

In the previous section, we argued that it is plausible to interpret HRG as providing an explanation by constraint. We believe there is also a second, distinct, non-causal reading of HRG's explanatoriness following the non-causal counterfactual account of explanation (Jansson and Saatsi 2019; Reutlinger 2016, 2018; Saatsi 2018b; Woodward 2018).

The non-causal counterfactual account extends Woodward's (2003) interventionist theory of causal explanation. According to it, causal explanations allow us to answer 'what-if-things-had-been-different' questions (what-if questions henceforth). They tell us what would happen to Y under a hypothetical intervention on X. The basic idea behind the non-causal counterfactual account is that although explanations always provide information about counterfactual dependence, that dependence need not be causal in the interventionist sense.

To illustrate, let us consider again the simple example above of Mother and her strawberries.[47] Jansson and Saatsi (2019; see also Woodward 2018) argue that we can interpret it along counterfactual lines. The counterfactual 'If Mother had had 24 strawberries, then she would have been able to divide them evenly between her 3 children' is true (cf. Lange 2017, 19–20). We can evaluate the truth of that counterfactual because we know that Mother's failure — or capacity — to divide evenly the strawberries *depends* on the number of strawberries she has. The facts that support this counterfactual relies on mathematics. They are not causal. We know that if Mother's *number* of strawberries had been adequately different, then she would have been able to divide them evenly. Crucially, the counterfactual does not depend on the particular physical details of the situation, but rather on mathematical facts.

Does HRG explain in virtue of citing non-causal difference-making relations between $\Delta\bar{g}$ and$r$, b, and c? We believe that so. But what makes a counterfactual explanation *causal*? Indeed, to show that HRG provides a non-causal counterfactual explanation, we can first ask whether it satisfies the criteria for causal explanation. If it does, this would indicate that our proposal is misguided. But if it does not, then this opens the door to a non-causal interpretation.

According to Woodward (Woodward 2018, 122; see also 2003), causal explanations have the following three features:

1. They provide answers to what-if questions that tell how Y changes,

[47] This does not imply that the two types of non-causal explanations are similar. Even though both accounts would consider this example to be of the non-causal sort, they reach that judgment for different reasons (see Reutlinger and Saatsi 2018). We remain agnostic concerning the relationship between these two types.

2. Under possible physical interventions on one or more (X_1, …,X_n), and

3. The relationship between X_1, …,X_n and Y is empirically (not conceptually) invariant under a range of interventions on X_1, …,$X_n$and background conditions.

Woodward suggests that we may have a non-causal explanation when we relax one or more of these conditions. The most important option involves keeping feature (1), but relaxing (2) or (3). In other words, non-causal explanations may allow answering what-if questions, but by appealing to non-causal counterfactual dependence. For instance, the case of Mother and her strawberries would violate (3) — the dependence is mathematical — and would satisfy (1) and (2) — we can intervene on the number of strawberries and it allows answering what-if questions. Let us look at these conditions in turn.

Does HRG provide answers to what-if questions? It seems to be a straightforward case of counterfactual dependence. The rule says that a change in the frequency of a trait depends on genetic relatedness as well as fitness costs and benefits. Since HRG is derived from substituting the Price equation into a regression equation, c and b are partial regression coefficients that relate changes between variables: had the value of b been different, then the value of $\Delta\bar{g}$would have been different. HRG thus allows us to answer what-if questions because of the counterfactual dependence relation between the variables. Should r, b and c be interpreted as causes of evolutionary change? In the case of HRG, we think it is more appropriate to consider that the counterfactual dependence is non-causal. This is because statistical association does not necessarily imply causation and therefore HRG might not express a causal relationship. The second step, then, is to look at whether or not the relation of counterfactual dependence that HRG expresses is causal.

Woodward's second criterion requires that it is possible to physically intervene on the explanatory variables.[48] For instance, one may explain the occurrence of bad weather by showing that had the atmospheric pressure been higher following an intervention, then the weather would have been fair. This relationship between atmospheric pressure and the weather is invariant under a range of interventions. However, intervening on the barometer needle would not change the weather. That we could hypothetically intervene on the atmospheric pressure and change the weather, but not on the barometer needle, shows that the two are not causally related — only atmospheric pressure is a cause of the weather.

---

[48] What a 'possible physical intervention' means has been a subject of debate in the literature (e.g. Reutlinger 2012). In a nutshell, the problem is that it seems the mere conceptual possibility of an intervention is not stringent enough, but that its actual realization is too demanding. Our discussion does not hinge on a specific solution to this problem.

The intervention criterion is at the centre stage of the discussion on HRG's causal interpretation. In short, the problem is to see whether or not HRG specifies what would happen upon hypothetical physical interventions (e.g. Birch 2017b, 72). As we indicated in section 2, the consensus so far in the literature is that there is no straightforward interpretation of HRG that satisfies the intervention criterion (2) (Birch 2017a; Nowak et al. 2017; Okasha 2016; Okasha and Martens 2016). In a recent analysis of HRG, Okasha and Martens (2016) compare the results of a hypothetical experiment to the partial regression definitions of costs and benefits. They demonstrate that these definitions have a causal meaning in the case of additive payoffs. The expected effect of an experimental intervention that randomly draws a selfish type from the population and switches it into an altruistic type corresponds to the partial regression coefficients of HRG. However, in the more general case of non-additive pairwise interactions, cost and benefit in HRG do not have causal meaning if the selfish types are picked at random from the population. The cost in HRG corresponds to the expected outcome of the hypothetical experiment only if the selfish types chosen to be experimentally manipulated are drawn from a specific cohort that meets a Fisherian condition of environmental constancy. This is not an innocuous assumption. As Okasha and Martens (2016) observe, this understanding of environmental constancy is particular to a simple evolutionary model of pairwise interaction with synergistic payoffs used to assess the cost of the social trait and a corresponding measure of assortment that lacks independent justification.

But even if we grant that HRG can receive a causal interpretation when additive effects hold, that actual social interactions are non-additive is the more realistic assumption (Grafen 2006, 543). And since the point of deriving HRG in the first place was to present a general and mathematically valid expression of HR (Queller 1992), limiting the scope of HRG to additive effects would require to restrict its intended domain of application. Hence, a causal interpretation of HRG along interventionist lines comes at a high cost.

If HRG cannot receive a causal interpretation in realistic evolutionary scenarios, then where does that leave us? HRG is not invariant under physical interventions because these may fail to bring about the expected changes in the case of non-additivity. Nevertheless, HRG exhibits a *statistical* relationship: it relates changes to the value of $\Delta\bar{g}$ to changes in the values of r, b or c. Moreover, that relationship is invariant in that it holds for all changes to the value of the variables. Not all statistical relationships are invariant in that sense. For instance, a statistical generalization represented in economics by the Phillips curve refers to the stable and inverse relationship between unemployment and inflation (or rate of nominal wage change). Due to stagflation in the 1970s, economists learned that this relationship was less invariant than they thought so. In contrast, HRG is highly invariant and this is one reason why it is so valuable as a statistical description of

the conditions under which social traits may evolve.[49] For example, suppose that r has a positive value and that b and c are such so that $rb > c$. Under these assumptions, altruism may evolve, viz. $\Delta\bar{g}$would be positive. However, had r been negative, we can infer that $rb > c$*could not* have been the case. The upshot is that HRG allows us to answer what-if questions concerning the different values $\Delta\bar{g}$ would take if r, b or c had been different. Therefore, if HRG explains, it is not in virtue of satisfying criterion (2). And if it does not satisfy that criterion, then the explanation would be non-causal.

Does this mean HRG also fails to meet condition (3), viz. that it holds for purely conceptual or mathematical reasons and not empirical ones? As we noted above, critics find fault with HRG's apparent tautological character (Allen, Nowak, and Wilson 2013; Nowak et al. 2017; Veelen et al. 2012, 2017). The worry is that HRG is trivially true and thus empirically empty. In particular, the issue is that since HRG does not assume anything about the dynamics that underlie r, b, or c, it is not possible to specify how the value of $\Delta\bar{g}$would change over time.

We agree with Birch (Birch 2017a, 2017b) that HRG is not in the business of providing particular causal explanations. For that purpose, one could employ models from population genetics or evolutionary game theory (Gardner, West, and Wild 2011, 1037). Using HRG instead amounts to what Weisberg (2013, sec. 7.1) calls 'generalized modelling.' HRG's target is not specific instances of, for example, altruism spreading in a population of eusocial insects. HRG's target is behavioral patterns that have features shared by all social behaviors, regardless of the detailed life histories of individuals that belong to a particular population. So it is not so much that HRG makes idealizations rather than it abstracts away from the particular details (Rubin 2018).[50] HRG states that the relationship between r, b, and c is invariant regardless of the causal details; the statistical relationship holds for any population.

That said, we nevertheless believe that HRG is not trivially true in the sense that it expresses an empirical relationship. Birch (Birch 2017a, 3; see also Veelen et al. 2012) provides an instructive analogy between HRG and the US presidential election. A candidate wins the election if and only if she wins more than 270 votes in the electoral college, which has 538 electors. At first sight, this does not appear to be very informative; it just states the conditions for winning the presidential election. The person who asks 'Who will win the presidential election?' and gets the following

---

[49] See Woodward (2003, sec. 6.4) for a discussion of degrees of invariance in the context of causal explanation.

[50] To be clear, the suggestion is not that HRG is a non-causal explanation *because* it is abstract. We agree with Reutlinger and Andersen (2016) that a high level of abstraction is not a good demarcation criterion between causal and non-causal explanations. For a related discussion, see Bokulich (Bokulich 2014, 2018).

answer 'The candidate who wins at least 270 votes in the electoral college' may be disappointed by the answer's lack of predictive value.

But this is simply not the type of questions the winning conditions — or HRG — are meant to answer. One acceptable answer to the question 'Why did Hillary Clinton lose the 2016 election?' would be that she failed to have the required number of votes. This may seem uninformative, but it still provides relevant and true information about the election.[51] For instance, it shows that winning the election depends on winning the electoral college, not on the general vote, which Clinton won. If you do not know how the US presidential election works, this is crucial information. And even though the state-by-state breakdown of the votes does not add any new information on the overall condition for winning (Birch 2017a, 3), it helps to answer more fine-grained explanatory questions. For instance, we could also answer the question about Clinton's defeat by citing the fact that she lost Florida and Pennsylvania. This is because the following counterfactual is true: 'If Clinton had won the states of Florida and Pennsylvania, then she would have won the presidential election.' Of course, it is possible to provide alternative explanations that appeal to different and presumably causal facts, e.g. the announcement in the days preceding the election by the FBI that they would reopen the investigation on Clinton's use of a private email server.

Even if one would accept that the US presidential election winning conditions may be explanatory, it can be argued that they do not express empirical truths. They are a constitutional convention akin to a conceptual truth. One way to understand HRG is that it expresses the 'winning conditions' for the evolution of social traits. For instance, it states that altruism may only be selected for if there are indirect fitness benefits. But unlike presidential elections, that HRG holds is an empirical matter. That natural selection operates via direct or indirect fitness is a hypothesis about the empirical world. One of the common characteristics of inclusive fitness models that provide testable predictions is the partition of fitness into direct or indirect components (e.g. Frank 1998; Rousset 2004). This carries over to HRG which partitions fitness using a multivariate regression equation. There is an important disanalogy between Mother and her strawberries or the US presidential election. No amount of empirical evidence could falsify the claim that Mother will never be able to divide evenly 23 by 3 or that a candidate may win the election with less than 270 votes. If, however, empirical research were to demonstrate that indirect fitness (or positive genetic relatedness) is not required for altruism to evolve, HRG could be rebutted, even though it is not directly tested.

We have argued that HRG allows to answer what-if questions and is empirical, but does not (always) support *physical* interventions. It thus satisfies the first and third of Woodward's (2018) criteria for causal explanation, but not the second. This suggests that HRG provides a non-causal counterfactual explanation. But as

---

[51] See Gardner (2020) for a similar argument concerning the Price equation.

Woodward (2018) observes, accepting non-interventionist counterfactuals creates an issue concerning explanatory relevance. The notion of intervention helps to capture explanatory asymmetry. To use a stock philosophical example, intervening on the position of the sun would make a difference to the length of the shadow cast by the flagpole. But the sun surely would not move if we were to intervene on the shadow. A cause explains its effect, not the other way around. It is the time-honoured problem of the asymmetry of explanation.

Prima facie, it seems HRG runs into this predicament. HRG expresses correlations and correlations are symmetric. If explanations are asymmetric, how could we explain with HRG? We want to explain $\Delta\bar{g}$by appealing to r, b, and c, but $\Delta\bar{g}$ does not explain these variables. However, HRG supports counterfactuals both ways: had $\Delta\bar{g}$ been different, then r would have been different too. How best to capture the asymmetry in non-causal explanations is an ongoing debate (e.g. Craver and Povich 2017; Khalifa, Millson, and Risjord 2018; Lange 2018a, 2019). Although it can be a problem for particular cases, we do not believe it poses a particular challenge in this one.

First of all, we should note that some non-causal explanations are symmetric in a similar derivational sense (Reutlinger 2018; Woodward 2018). Thus, the fact that the counterfactual dependence is symmetric is not sufficient ground to rule out explanatory asymmetry. What it tells us is that counterfactual dependence by itself does not entail explanatory relevance, a conclusion we do not want to resist (see Pincock 2018). But then, what could be the source of explanatory asymmetry?

We are sympathetic to Lange's (2019) suggestion that there may not be sufficient and necessary conditions for explanatory asymmetry. Hence, the criteria that apply in one case may not be adequate in another. In fact, we believe that in that respect HR is a peculiar case. Initially, Hamilton's motivation was to find a causal explanation for altruism (Frank 2013). His original analysis emphasized the causal decomposition of total fitness effects into the relatedness, costs, and benefits of social action. Following Hamilton, many theoretical biologists refined this analysis and developed methods that connect HR to genetics and evolutionary game theory. In this process, Queller (1992) derived a general regression version of the rule that is valid in cases where there is a complicated relationship between genotype, phenotype, and fitness. But, ultimately, what indirectly vindicates HRG is the empirical success of predictions derived from models that apply methods from population genetics and evolutionary game theory. HRG is just a generalization of these models (Queller 1992, 377) and starting from HRG, one can mathematically derive the main result of particular models that make empirical predictions (Gardner, West, and Wild 2011). Moreover, it is those models with mechanistic details that provide the direction of explanation. Fitness effects and relatedness explain whether a trait is selected and not vice versa because this is what we model and observe empirically.

Now, would HRG explain in a world with only HRG, but without the other less general versions of the rule and the empirical evidence that supports them? No. But

the fact that less general models work give us reasons to believe that the more general and non-causal one, HRG, also works. So although the justification of HRG comes from more detailed causal models, it does not explain by virtue of identifying causes. HRG would be devoid of any empirical meaning or justification if not for its causal counterparts.[52] This is why, in practice, biologists cannot avoid using less general models based on HRS and HRA (Bourke 2014; Gardner, West, and Wild 2011). But being able to show that the relation described by HRG is invariant also strengthens the justification for the causal models. Models have to explore mechanisms related to one of the two evolutionary pathways that affect inclusive fitness and select for social traits. For instance, the invariance makes it extremely unlikely that we would ever find altruism without mechanisms that generate sufficiently high indirect fitness benefits. This is true regardless of the particular causal details. And that we do not find altruism without those mechanisms reinforces the case for HRG. There is thus a sort of feedback relationship between the causal and non-causal versions of HR. In that sense, HRG and the more detailed causal models complement each other (Andersen 2018).

To sum up, HRG would satisfy criteria (1) — answering what-if questions — and (3) — invariant for empirical reasons —, but not (2). If HRG explains and does so in virtue of identifying a relation of counterfactual dependence, but does not satisfy (2), then it provides a non-causal counterfactual explanation.[53] Of courses, this assumes that intervention is necessary for causation. This assumption is sometimes disputed (Reutlinger 2012; Strevens 2008). We have focused on the interventionist framework because it is the one HRG has been assessed against in the literature. Other theories of causal explanation may thus reach a different verdict. Our goal is not to settle this question. If it turns out that HRG can receive a plausible causal interpretation, so be it. Our aim was more modest, viz. to show that HRG can explain *despite* a lack of causal interpretation.

6. Conclusion

We have argued that even though HRG may serve as an organizing framework, it does not explain for *that* reason. Since we share the idea that HRG does not explain in virtue of citing causes, we have proposed to interpret HRG as providing a non-

---

[52] This is also why we could not simply choose any predictor of fitness and calculate HRG, e.g. the moon phase (see Queller 2011), even though it would lead to a mathematically valid version of the rule. The moon phase is not causally related to fitness and the evolution of social traits.

[53] Cases with a similar structure have been discussed in the literature (e.g. Bokulich 2011, 2018; Rice 2015; Saatsi 2018c; Saatsi and Pexton 2013; Reutlinger 2016).

causal explanation according to two other leading accounts. More precisely, we maintain that HRG can either be viewed as providing an explanation by constraint or a non-causal counterfactual explanation. These accounts better identify the reasons why HRG explains the evolution of social traits.

References

Abbot, Patrick, Jun Abe, John Alcock, Samuel Alizon, Joao A. C. Alpedrinha, Malte Andersson, Jean-Baptiste Andre, et al. 2011. "Inclusive Fitness Theory and Eusociality." *Nature* 471 (7339): E1–4. https://doi.org/10.1038/nature09831.

Allen, Benjamin, Martin A. Nowak, and Edward O. Wilson. 2013. "Limitations of Inclusive Fitness." *Proceedings of the National Academy of Sciences* 110 (50): 20135–39. https://doi.org/10.1073/pnas.1317588110.

Andersen, Holly. 2018. "Complements, Not Competitors: Causal and Mathematical Explanations." *The British Journal for the Philosophy of Science* 69 (2): 485–508. https://doi.org/10.1093/bjps/axw023.

Baron, Sam. 2019. "Mathematical Explanation by Law." *The British Journal for the Philosophy of Science* 70 (3): 683–717. https://doi.org/10.1093/bjps/axx062.

Baron, Sam, Mark Colyvan, and David Ripley. 2017. "How Mathematics Can Make a Difference." *Philosophers' Imprint* 17 (3): 1–19.

Birch, Jonathan. 2014. "Hamilton's Rule and Its Discontents." *The British Journal for the Philosophy of Science* 65 (2): 381–411. https://doi.org/10.1093/bjps/axt016.

———. 2017a. "The Inclusive Fitness Controversy: Finding a Way Forward." *Royal Society Open Science* 4 (7): 170335. https://doi.org/10.1098/rsos.170335.

———. 2017b. *The Philosophy of Social Evolution*. Oxford: Oxford University Press.

Birch, Jonathan, and Samir Okasha. 2015. "Kin Selection and Its Critics." *BioScience* 65 (1): 22–32. https://doi.org/10.1093/biosci/biu196.

Bokulich, Alisa. 2011. "How Scientific Models Can Explain." *Synthese* 180 (1): 33–45.

———. 2014. "How the Tiger Bush Got Its Stripes: 'How Possibly' Vs. 'How Actually' Model Explanations." *The Monist* 97 (3): 321–38. https://doi.org/10.5840/monist201497321.

———. 2018. "Searching for Non-Causal Explanations in a Sea of Causes." In *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, edited by Alexander Reutlinger and JuhaSaatsi, 141–63. Oxford: Oxford University Press.

Bourke, Andrew F. G. 2014. “Hamilton’s Rule and the Causes of Social Evolution.” *Phil. Trans. R. Soc. B* 369 (1642): 20130362. https://doi.org/10.1098/rstb.2013.0362.

Bowles, Samuel, and Herbert Gintis. 2004. “The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations.” *Theoretical Population Biology* 65 (1): 17–28. https://doi.org/10.1016/j.tpb.2003.07.001.

Cavalli-Sforza, L. L., and M. W. Feldman. 1978. “Darwinian Selection and ‘Altruism’.” *Theoretical Population Biology* 14 (2): 268–80. https://doi.org/10.1016/0040-5809(78)90028-X.

Craver, Carl F., and Mark Povich. 2017. “The Directionality of Distinctively Mathematical Explanations.” *Studies in History and Philosophy of Science Part A* 63: 31–38. https://doi.org/10.1016/j.shpsa.2017.04.005.

Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection*. London: John Murray.

———. 1871. *The Descent of Man, and Selection in Relation to Sex*. New York: D. Appleton & Company.

De Regt, Henk W. 2009. “The Epistemic Value of Understanding.” *Philosophy of Science* 76 (5): 585–97.

———. 2017. *Understanding Scientific Understanding*. New York: Oxford University Press.

De Regt, Henk W., and Dennis Dieks. 2005. “A Contextual Approach to Scientific Understanding.” *Synthese* 144 (1): 137–70.

Dellsén, Finnur. 2020. “Beyond Explanation: Understanding as Dependency Modelling.” *The British Journal for the Philosophy of Science* 71 (4): 1261–86. https://doi.org/10.1093/bjps/axy058.

Elgin, Catherine Z. 2017. *True Enough*. Cambridge, MA: MIT Press.

Elgin, Mehmet, and Elliott Sober. 2015. “Causal, A Priori True, and Explanatory: A Reply to Lange and Rosenberg.” *Australasian Journal of Philosophy* 93 (1): 167–71. https://doi.org/10.1080/00048402.2014.948022.

Frank, Steven A. 1998. *Foundations of Social Evolution*. Princeton, NJ: Princeton University Press.

———. 2012. “Natural Selection. IV. The Price Equation*.” *Journal of Evolutionary Biology* 25 (6): 1002–19. https://doi.org/10.1111/j.1420-9101.2012.02498.x.

———. 2013. “Natural Selection. VII. History and Interpretation of Kin Selection Theory.” *Journal of Evolutionary Biology* 26 (6): 1151–84. https://doi.org/10.1111/jeb.12131.

Frank, Steven A., and Gordon A. Fox. 2020. “The Inductive Theory of Natural Selection.” In *The Theory of Evolution: Principles, Concepts, and Assumptions*, edited by Samuel M. Scheiner and David P. Mindell, 171–93. Chicago; London: University of Chicago Press.

Gardner, Andy. 2015. “Hamilton’s Rule.” *The American Naturalist* 186 (2): ii–iii. https://doi.org/10.1086/682082.

———. 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.

Gardner, Andy, and Stuart A. West. 2014. “Inclusive Fitness: 50 Years On.” *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 369 (1642): 20130356. https://doi.org/10.1098/rstb.2013.0356.

Gardner, Andy, Stuart A. West, and Nicholas H. Barton. 2007. “The Relation Between Multilocus Population Genetics and Social Evolution Theory.” *The American Naturalist* 169 (2): 207–26. https://doi.org/10.1086/510602.

Gardner, Andy, Stuart A. West, and Geoff Wild. 2011. “The Genetical Theory of Kin Selection.” *Journal of Evolutionary Biology* 24 (5): 1020–43. https://doi.org/10.1111/j.1420-9101.2011.02236.x.

Gijsbers, Victor. 2013. “Understanding, Explanation, and Unification.” *Studies in History and Philosophy of Science Part A* 44 (3): 516–22. https://doi.org/10.1016/j.shpsa.2012.12.003.

———. 2014. “Unification as a Measure of Natural Classification.” *THEORIA. An International Journal for Theory, History and Foundations of Science* 29 (1): 71. https://doi.org/10.1387/theoria.7235.

Gintis, Herbert. 2000. “Strong Reciprocity and Human Sociality.” *Journal of Theoretical Biology* 206 (2): 169–79. https://doi.org/10.1006/jtbi.2000.2111.

Grafen, Alan. 2006. “Optimization of Inclusive Fitness.” *Journal of Theoretical Biology* 238 (3): 541–63. https://doi.org/10.1016/j.jtbi.2005.06.009.

Hamilton, W. D. 1963. “The Evolution of Altruistic Behavior.” *The American Naturalist* 97 (896): 354–56. https://doi.org/10.1086/497114.

———. 1964a. “The Genetical Evolution of Social Behavior. I.” *Journal of Theoretical Biology* 7 (1): 1–16. https://doi.org/10.1016/0022-5193(64)90038-4.

———. 1964b. “The Genetical Evolution of Social Behaviour. II.” *Journal of Theoretical Biology* 7 (1): 17–52. https://doi.org/10.1016/0022-5193(64)90039-6.

———. 1970. “Selfish and Spiteful Behaviour in an Evolutionary Model.” *Nature* 228 (5277): 1218–20. https://doi.org/10.1038/2281218a0.

Jansson, Lina, and JuhaSaatsi. 2019. “Explanatory Abstractions.” *The British Journal for the Philosophy of Science* 70 (3): 817–44. https://doi.org/10.1093/bjps/axx016.

Karlin, Samuel, C. Matessi, Walter Fred Bodmer, and John Frank Charles Kingman. 1983. “The Eleventh R. A. Fisher Memorial Lecture - Kin Selection and Altruism.” *Proceedings of the Royal Society of London. Series B. Biological Sciences* 219 (1216): 327–53. https://doi.org/10.1098/rspb.1983.0077.

Kelp, Christoph. 2015. “Understanding Phenomena.” *Synthese* 192 (12): 3799–3816. https://doi.org/10.1007/s11229-014-0616-x.

Khalifa, Kareem, Jared Millson, and Mark Risjord. 2018. “Inference, Explanation, and Asymmetry.” *Synthese*, 1–25. https://doi.org/10.1007/s11229-018-1791-y.

Kitcher, Philip. 1981. “Explanatory Unification.” *Philosophy of Science* 48 (4): 507–31.

———. 1989. “Explanatory Unification and the Causal Structure of the World.” In *Scientific Explanation*, edited by Philip Kitcher and Wesley C. Salmon, 410–505. Minneapolis: University of Minnesota Press.

Lange, Marc. 2009. *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*. Oxford: Oxford University Press.

———. 2013. “What Makes a Scientific Explanation Distinctively Mathematical?” *The British Journal for the Philosophy of Science* 64 (3): 485–511. https://doi.org/10.1093/bjps/axs012.

———. 2017. *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. New York: Oxford University Press.

———. 2018a. “A Reply to Craver and Povich on the Directionality of Distinctively Mathematical Explanations.” *Studies in History and Philosophy of Science Part A*, January. https://doi.org/10.1016/j.shpsa.2018.01.002.

———. 2018b. “Because Without Cause: Scientific Explanations by Constraint.” In *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, edited by Alexander Reutlinger and JuhaSaatsi, 15–38. Oxford: Oxford University Press.

———. 2019. “Asymmetry as a Challenge to Counterfactual Accounts of Non-Causal Explanation.” *Synthese*. https://doi.org/10.1007/s11229-019-02317-3.

Lange, Marc, and Alexander Rosenberg. 2011. “Can There Be A Priori Causal Models of Natural Selection?” *Australasian Journal of Philosophy* 89 (4): 591–99. https://doi.org/10.1080/00048402.2011.598175.

Lehmann, Laurent, and François Rousset. 2014. “The Genetical Theory of Social Behaviour.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1642): 20130357. https://doi.org/10.1098/rstb.2013.0357.

Levin, Samuel R., and Alan Grafen. 2019. “Inclusive Fitness Is an Indispensable Approximation for Understanding Organismal Design.” *Evolution* 73 (6): 1066–76. https://doi.org/10.1111/evo.13739.

Lipton, Peter. 2009. “Understanding Without Explanation.” In *Scientific Understanding. Philosophical Perspectives*, edited by Henk W. De Regt, Sabina Leonelli, and Kai Eigner, 43–63. Pittsburgh: University of Pittsburgh Press.

Luque, Victor 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.

Mancosu, Paolo. 2001. “Mathematical Explanation: Problems and Prospects.” *Topoi* 20 (1): 97–117. https://doi.org/10.1023/A:1010621314372.

Matessi, Carlo, and Samuel Karlin. 1984. “On the Evolution of Altruism by Kin Selection.” *Proceedings of the National Academy of Sciences* 81 (6): 1754–58. https://doi.org/10.1073/pnas.81.6.1754.

Newman, Mark. 2017. “Theoretical Understanding in Science.” *British Journal for the Philosophy of Science* 68 (2): 571–95. https://doi.org/10.1093/bjps/axv041.

Nowak, Martin A., Alex McAvoy, Benjamin Allen, and Edward O. Wilson. 2017. “The General Form of Hamilton’s Rule Makes No Predictions and Cannot Be Tested Empirically.” *Proceedings of the National Academy of Sciences* 114 (22): 5665–70. https://doi.org/10.1073/pnas.1701805114.

Nowak, Martin A., Corina E. Tarnita, and Edward O. Wilson. 2010. “The Evolution of Eusociality.” *Nature* 466 (7310): 1057–62. https://doi.org/10.1038/nature09205.

Okasha, Samir. 2016. “On Hamilton’s Rule and Inclusive Fitness Theory with Nonadditive Payoffs.” *Philosophy of Science* 83 (5): 873–83. https://doi.org/10.1086/687871.

Okasha, Samir, and Johannes Martens. 2016. “The Causal Meaning of Hamilton’s Rule.” *Royal Society Open Science* 3 (3): 160037. https://doi.org/10.1098/rsos.160037.

Pincock, Christopher. 2015. “Abstract Explanations in Science.” *The British Journal for the Philosophy of Science* 66 (4): 857–82. https://doi.org/10.1093/bjps/axu016.

———. 2018. “Accommodating Explanatory Pluralism.” In *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, edited by Alexander Reutlinger and JuhaSaatsi, 39–56. Oxford: Oxford University Press.

Price, George R. 1970. “Selection and Covariance.” *Nature* 227 (5257): 520–21. https://doi.org/10.1038/227520a0.

———. 1972. “Extension of Covariance Selection Mathematics.” *Annals of Human Genetics* 35 (4): 485–90. https://doi.org/10.1111/j.1469-1809.1957.tb01874.x.

Queller, David C. 1984. “Kin Selection and Frequency Dependence: A Game Theoretic Approach.” *Biological Journal of the Linnean Society* 23 (2-3): 133–43. https://doi.org/10.1111/j.1095-8312.1984.tb00133.x.

———. 1985. “Kinship, Reciprocity and Synergism in the Evolution of Social Behaviour.” *Nature* 318 (6044): 366. https://doi.org/10.1038/318366a0.

———. 1992. “A General Model for Kin Selection.” *Evolution* 46 (2): 376–80. https://doi.org/10.1111/j.1558-5646.1992.tb02045.x.

———. 2011. “Expanded Social Fitness and Hamilton’s Rule for Kin, Kith, and Kind.” *Proceedings of the National Academy of Sciences* 108 (Supplement 2): 10792–99. https://doi.org/10.1073/pnas.1100298108.

Räz, Tim. 2017. “The Silent Hexagon: Explaining Comb Structures.” *Synthese* 194 (5): 1703–24. https://doi.org/10.1007/s11229-016-1014-3.

Reutlinger, Alexander. 2012. “Getting Rid of Interventions.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, Causality in the Biomedical and Social Sciences, 43 (4): 787–95. https://doi.org/10.1016/j.shpsc.2012.05.006.

———. 2016. “Is There A Monist Theory of Causal and Noncausal Explanations? The Counterfactual Theory of Scientific Explanation.” *Philosophy of Science* 83 (5): 733–45. https://doi.org/10.1086/687859.

———. 2017. “Explanation Beyond Causation? New Directions in the Philosophy of Scientific Explanation.” *Philosophy Compass* 12 (2): 1–11. https://doi.org/10.1111/phc3.12395.

———. 2018. “Extending the Counterfactual Theory of Explanation.” In *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, edited by Alexander Reutlinger and JuhaSaatsi, 74–95. Oxford: Oxford University Press.

Reutlinger, Alexander, and Holly Andersen. 2016. “Abstract Versus Causal Explanations?” *International Studies in the Philosophy of Science* 30 (2): 129–46. https://doi.org/10.1080/02698595.2016.1265867.

Reutlinger, Alexander, and JuhaSaatsi. 2018. “Introduction.” In *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, edited by Alexander Reutlinger and JuhaSaatsi, 1–12. Oxford: Oxford University Press.

Rice, Collin. 2015. “Moving Beyond Causes: Optimality Models and Scientific Explanation.” *Noûs* 49 (3): 589–615. https://doi.org/10.1111/nous.12042.

———. 2016. “Factive Scientific Understanding Without Accurate Representation.” *Biology & Philosophy* 31 (1): 81–102. https://doi.org/10.1007/s10539-015-9510-2.

Rohwer, Yasha, and Collin Rice. 2013. “Hypothetical Pattern Idealization and Explanatory Models.” *Philosophy of Science* 80 (3): 334–55. https://doi.org/10.1086/671399.

Rousset, François. 2004. *Genetic Structure and Selection in Subdivided Populations*. Princeton, NJ: Princeton University Press.

———. 2015. “Regression, Least Squares, and the General Version of Inclusive Fitness.” *Evolution* 69 (11): 2963–70. https://doi.org/10.1111/evo.12791.

Rubin, Hannah. 2018. “The Debate over Inclusive Fitness as a Debate over Methodologies.” *Philosophy of Science* 85 (1): 1–30. https://doi.org/10.1086/694809.

Saatsi, Juha. 2012. “Mathematics and Program Explanations.” *Australasian Journal of Philosophy* 90 (3): 579–84. https://doi.org/10.1080/00048402.2012.665374.

———. 2018a. “A Pluralist Account of Non-Causal Explanation in Science and Mathematics.” *Metascience* 27 (1): 3–9. https://doi.org/10.1007/s11016-017-0249-z.

———. 2018b. “On Explanations from Geometry of Motion.” *The British Journal for the Philosophy of Science* 69 (1): 253–73. https://doi.org/10.1093/bjps/axw007.

———. 2018c. “Realism and the Limits of Explanatory Reasoning.” In *The Routledge Handbook of Scientific Realism*, edited by JuhaSaatsi, 200–211. Milton Park: Routledge.

Saatsi, Juha, and Mark Pexton. 2013. “Reassessing Woodward’s Account of Explanation: Regularities, Counterfactuals, and Noncausal Explanations.” *Philosophy of Science* 80 (5): 613–24. https://doi.org/10.1086/673899.

Sober, Elliott. 2011. “A Priori Causal Models of Natural Selection.” *Australasian Journal of Philosophy* 89 (4): 571–89. https://doi.org/10.1080/00048402.2010.535006.

Steiner, Mark. 1978. “Mathematical Explanation.” *Philosophical Studies* 34 (2): 135–51. https://doi.org/10.1007/BF00354494.

Strevens, Michael. 2008. *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press.

———. 2013. “No Understanding Without Explanation.” *Studies in History and Philosophy of Science Part A* 44 (3): 510–15. https://doi.org/10.1016/j.shpsa.2012.12.005.

Taylor, Peter D., and Steven A. Frank. 1996. “How to Make a Kin Selection Model.” *Journal of Theoretical Biology* 180 (1): 27–37. https://doi.org/10.1006/jtbi.1996.0075.

Veelen, Matthijs van. 2009. “Group Selection, Kin Selection, Altruism and Cooperation: When Inclusive Fitness Is Right and When It Can Be Wrong.” *Journal of Theoretical Biology* 259 (3): 589–600. https://doi.org/10.1016/j.jtbi.2009.04.019.

Veelen, Matthijs van, Benjamin Allen, Moshe Hoffman, Burton Simon, and Carl Veller. 2017. “Hamilton’s Rule.” *Journal of Theoretical Biology* 414 (February): 176–230. https://doi.org/10.1016/j.jtbi.2016.08.019.

Veelen, Matthijs van, Julián García, Maurice W. Sabelis, and Martijn Egas. 2012. “Group Selection and Inclusive Fitness Are Not Equivalent; the Price Equation Vs. Models and Statistics.” *Journal of Theoretical Biology*, Evolution of Cooperation, 299 (April): 64–80. https://doi.org/10.1016/j.jtbi.2011.07.025.

Vromen, Jack. 2012. “Human Cooperation and Reciprocity.” In *Evolution and Rationality: Decisions, Co-Operation and Strategic Behaviour*, edited by Samir Okasha and Ken Binmore, 158–84. Cambridge: Cambridge University Press.

Weisberg, Michael. 2013. *Simulation and Similarity. Using Models to Understand the World*. Oxford: Oxford University Press.

West, Stuart A., Claire El Mouden, and Andy Gardner. 2011. “Sixteen Common Misconceptions about the Evolution of Cooperation in Humans.” *Evolution and Human Behavior* 32 (4): 231–62. https://doi.org/10.1016/j.evolhumbehav.2010.08.001.

Woodward, James. 2003. *Making Things Happen. A Theory of Causal Explanation*. New York: Oxford University Press.

———. 2018. “Some Varieties of Non-Causal Explanation.” In *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, edited by Alexander Reutlinger and JuhaSaatsi, 117–38. Oxford: Oxford University Press.

Ylikoski, Petri, and N. EmrahAydinonat. 2014. “Understanding with Theoretical Models.” *Journal of Economic Methodology* 21 (1): 19–36. https://doi.org/10.1080/1350178X.2014.886470.

Chapter 6:Conclusion

There is significant value of scientific research on the intersection of economics and biology. First, knowledge about the evolutionary foundations of economic choices and outcomes improves our understanding of the causes and consequences of individual behaviour. Second, information about evolved economic traits may result in targeted interventions to prevent outcomes detrimental to society. It is reasonable to assume that the interest in these issues will only grow in the upcoming years due to growth in the collection of genetic and behavioural data.

To conclude this thesis, it is important to define a way forward and discuss several potential avenues for future research, a number of research questions and ideas open to further investigation. In relation to the topics discussed in this thesis, future research could address three main issues.

- First, provide a detailed description of proximate mechanisms and identify how they respond in a range of environments to produce behavior
- Second, investigate the evolutionary success of different proximate mechanisms from a theoretical perspective and predict which ones could evolve
- Third, examine the mathematical structure of evolutionary models and see what kind of explanation (i.e. mathematical, causal, non-causal etc) they provide.

Among the more exciting new developments in economics is a stream of studies that consider the interplay of cultural factors in the explanation of human traits. At the level of decision making, researchers can examine social learning biases in more detail before evolutionary researchers begin to understand why they have evolved. A distinction between those proximate mechanisms that are fixed and those that are flexible within an individual's lifetime is crucial both for experimental and theoretical research. Another key issue is whether we have evolved psychological mechanisms that work well only in particular environments or across a range of different environments.

It is also important to clarify the key role of evolutionary theory in future study of human traits. Theoretical economics and biology place a large emphasis on formal analytical models and computer simulation. Although it is generally preferable to have mathematical models which can be solved analytically, adding important details can lead to intractable analytical solutions. With respect to evolutionary models, it is important to develop theoretical tools that allow a more systematic approach to predicting which types of proximate mechanisms will be favored by natural selection and cultural evolution. One key issue here is whether the widespread adoption of particular proximate mechanisms by individuals in a

population prevents alternative proximate mechanisms to invade. One reason that such theoretical effort is likely to succeed is the recent advances in the indirect evolutionary approach and cognitive ecology and their application to wide range of issues related to human and animal behavior.

Finally, there is a cluster of topics in this thesis that will appeal to philosophers of social science. Both in economics and in biology, there are types of models that provide ultimate explanations that have unique characteristics. A recent line of theoretical research does not examine causality but focuses on statistical correlates. Establishing causation is often difficult in biology and the social sciences and the study of correlations is usually a first step toward further causal explanations. What is important to study is how evolutionary models explain concrete phenomena. For example, the Price equation is often used to generate evolutionary models in economics and biology and these models ignore causal relations and mechanistic details. Philosophical research could analyze particular accounts of non-causal and causal explanation and examine the kind of explanation these evolutionary models provide.

Summary (English)

Social sciences and biology have been interacting throughout their histories. Multiple disciplines have applied Darwin's theory, including psychology and economics. Within economics, there has been a resurgence of interest in using evolution theory for the study of economic traits, similar to the way that it is applied in the biological sciences. This research is typically organized around Ernst Mayr's ultimate-proximate distinction. Ultimate explanations of behaviour invoke evolutionary causes (i.e. natural selection) to answer why-questions (why did a particular behaviour evolve compared to other behaviours that could have evolved in the lineage?). Proximate explanations of behaviour typically invoke behaviour-generating causes (e.g. psychological, neurological mechanisms) inside organisms and environmental conditions to answer how-questions (how is the behaviour produced within the organism?).

The proximate-ultimate distinction has been the cornerstone of evolutionary research. It is possible, however, to raise a different research question when it comes to explaining human evolution. What is important to observe is that behaviour is brought about by psychological mechanisms, at an intermediate level between genes and behaviour. Hence, we can go beyond Mayr's distinction and pose the following question: Why did a particular proximate mechanism evolve rather than some other that could have produced a set of behaviours? In this dissertation I discuss how the mechanism selection question fits into current debates in biology, economics and philosophy.

The Darwinian theory of evolution has arguably become an important building block for experimental and theoretical economists. According to Burnham, it is possible to formulate novel hypotheses and predictions about human preferences, on the basis of what patterns of behavior would have been adaptive in the ancestral environment. After clarifying two theoretical concepts, the Adaptively Relevant Environment and fitness maximization, in chapter 2 I argue that multiple scientifically plausible hypotheses about human preferences are compatible with evolutionary models that target behavior. Moreover, I propose a refinement of Burnham's method based on theoretical resources provided by the indirect evolutionary approach. Economists apply or build evolutionary models of their own that target particular mechanisms of human psychology and cognition. Such models may reduce the number of plausible hypotheses to allow for rigorous scientific testing in laboratory or field experiments.

Recently a number of papers draw upon ideas from sexual selection and costly signalling theory to argue that conspicuous consumption has evolved as a sexually selected mating strategy that provides reproductive benefits to men. In chapter 3, I outline what are considered to be the criteria for arguing that a trait is the outcome of sexual selection and I explore whether conspicuous consumption does in fact possess the main features of sexual adaptations. Though I share the insight that

evolutionary theory can contribute to the explanation of consumption practices, I argue that evolutionary researchers attempt to examine the evolutionary underpinnings of conspicuous consumption without analyzing in detail human evolved psychology and the available evidence about the ancestral environment. I further argue that cultural evolution theory provides an alternative explanation of conspicuous consumption in modern environments. In particular, conspicuous consumption is understood as a pattern of behavior marked by specific social learning mechanisms. Such an approach reflects the analytical tools of cultural evolution theory and provides a classification of motivational factors involved in consumption practices.

Work in evolutionary science devotes specific attention to non-verbal expressions of emotions. In chapter 4, I examine Robert Frank's influential attempt to explain emotions, non-verbal markers and social behaviors. Following this line of work, I focus on the green-beard explanation of social behaviors. In response to the criticisms raised against this controversial ultimate explanation, based on resources found in Frank's work, I propose an alternative red-beard explanation of human sociality. The red-beard explanation explains the emergence and evolution of emotions, a proximate cause, rather than patterns of behavior. In contrast to simple evolutionary models that invoke a green-beard mechanism, I demonstrate that the red-beard explanation can be evolutionary stable. Social emotions are a common cause of a social behavior and a phenotypic marker and therefore cooperative behavior cannot be suppressed without also changing the marker.

Finally, the explanatory power of Hamilton's rule, the main explanatory principle of social evolution theory, is an ongoing subject of controversy. In chapter 5, Philippe Verreault-Julien and I reinforce the case for the considerable value of the regression-based version of the rule in explaining the evolution of social traits. Although we agree that the rule can have an organizing role in social evolution research, we maintain that it does not explain in virtue of citing mechanism and causes or providing an organizing framework. Instead, we argue it either provides an explanation by constraint or a non-causal counterfactual explanation

Summary (Dutch)

Door de geschiedenis heen hebben zich verschillende kruisbestuivingen voorgedaan tussen de sociale en biologische wetenschappen. Darwin's theorie bijvoorbeeld is in vele wetenschapsgebieden toegepast, waaronder de psychologische en de economische. Met name binnen deze laatstgenoemde discipline wordt evolutietheorie opnieuw bestudeert vanwege de studie naar economische gedragseigenschappen op soortgelijke wijze waarop deze wordt toegepast in de biologie. Dit onderzoek is in hoofdlijnen georganiseerd rond Ernst Mayr's onderscheid tussen ultieme en nabije gedragsverklaringen (proximate-ultimate explanations of behavior). Ultieme gedragsverklaringen maken gebruik van evolutionaire oorzakelijkheden (zoals natuurlijke selectie) om zogenaamde waarom-vragen te beantwoorden: Waarom evolueerde bepaald gedrag in plaats van andere erfelijke gedragsopties? Anderzijds impliceren nabije gedragsverklaringen gedragsgenererende oorzaken, zoals psycho- en neurologische mechanismen, om hoe-vragen te beantwoorden.

Het ultieme-nabije onderscheid is een van de hoekstenen van evolutionair onderzoek. Als verklaring voor menselijke ontwikkeling in het algemeen is het mogelijk om een andere onderzoeksvraag te berde te brengen dan welke aan dit onderscheid ten grondslag ligt. Het is belangrijk te observeren dat gedrag ook wordt voortgebracht door psychologische processen en wel op het bemiddelend niveau tussen genen en gedrag. Daarom is het mogelijk voorbij te gaan aan Mayr's onderscheid door de volgende vraag te stellen: waarom evolueerde het ene nabijheidsmechanisme in plaats van het andere? In deze dissertatie bespreek ik hoe de vraag naar dit selectiemechanisme past bij recente debatten binnen de biologie, economie en filosofie.

De darwiniaanse evolutietheorie vormt daarnaast een belangrijke bouwsteen voor experimentele en theoretische economen. Volgens Burnham is het mogelijk om nieuwe hypothesen en voorspellingen over menselijke voorkeuren te vormen op basis van gedragspatronen die adaptief zouden zijn in de voorouderlijke leefomgeving. In deze dissertatie licht ik allereerst twee theoretische concepten toe: de Adaptatie Ontvankelijke Leefomgeving en fitness-maximalisatie. In het tweede hoofdstuk beargumenteer ik dat wat betreft menselijke voorkeuren, er meer dan één wetenschappelijk plausibele hypothese verenigbaar is met evolutionaire gedragsmodellen. Bovendien doe ik een voorstel om Burnham's methode bij te stellen op basis van de theoretische bronnen die geboden worden door de indirecte benadering van de evolutietheorie. De door economen gemaakte of toegepaste evolutiemodellen hebben tot doel bepaalde wordingsprocessen van menselijke psychologie en cognitie. Dergelijke modellen zouden de hoeveelheid plausibele hypothesen kunnen beperken en daarmee de haalbaarheid van grondig empirisch en wetenschappelijk lab- of veldonderzoek met informatieve resultaten.

Een aantal recente papers beargumenteren, mede geïnspireerd door seksuele selectie en signaleringstheorie, dat opzichtige consumptie geëvolueerd is als een seksueel geselecteerde paringsstrategie dat mannen reproductievoordelen oplevert. Het derde hoofdstuk zet de gangbare argumentatiecriteria uiteen waarom een gedragsbepaling (trait) het resultaat is van seksuele selectie en verken ik of de mogelijkheid dat opzichtige consumptie inderdaad de hoofdkenmerken bezit van seksuele adaptaties. Ondanks dat ik het inzicht onderschrijf dat evolutietheorie kan bijdragen aan verklaringen van consumptie, beargumenteer ik dat de onderzoekspogingen zowel de evolutionaire psychologie van de mens, als enkele bewijzen uit de voorouderlijke leefomgeving ofwel miskennen of niet voldoende verdisconteren. Vervolgens laat ik zien dat culturele evolutietheorie een andere uitleg geeft aan opzichtige consumptie in de moderne tijd. In het bijzonder omdat opzichtige consumptie begrepen wordt als een gedragspatroon dat wordt gekenmerkt door specifieke sociale leer- en omgangsvormen. Een dergelijke aanpak verwijst naar de analytische middelen uit de culture evolutietheorie en voorziet in een classificatie van beweegredenen binnen consumptiepraktijken.

Evolutionair-wetenschappelijk werk besteedt in het bijzonder aandacht aan non-verbale uitdrukkingen van emoties. Daarom onderzoek ik in het vierde hoofdstuk Robert Frank's algemeen bekende pogingen om emoties en non-verbale uitdrukkingen van sociaal gedrag te verklaren. In navolging van dit werk, leg ik de nadruk op het zogenoemde groene baard-mechanisme van sociale gedragingen. Als antwoord op de kritiek die naar voren is gebracht tegen deze controversiële ultieme gedragsverklaring en gebaseerd op bronnen die men ook in Frank's werk kan vinden, stel ik een alternatief voor: de rode baard-verklaring van menselijke socialiteit. In plaats van gedragspatronen legt de rode baard-verklaring het ontstaan en evolutie uit van emoties, een nabije oorzaak die er immers mede aan ten grondslag liggen. In tegenstelling tot eenvoudige evolutiemodellen die gebruik maken van het groene baard-mechanisme, laat ik zien dat de rode baard-verklaring evolutionair stabiel kan zijn. Sociale emoties zijn een veelvoorkomende oorzaak van een sociale gedraging en een fenotypisch uitdrukking en dus kan coöperatief gedrag niet worden onderdrukt zonder dat deze uitdrukking verandert.

Als laatste bespreek ik de zeggingskracht van Hamilton's regel: als het algemeen uitleggend principe van sociale evolutietheorie en als controverse. In het vijfde hoofdstuk breek ik samen met Philippe Verreault-Julien een lans voor de aanzienlijke waarde van de regressie-versie van de regel om de evolutie van sociale gedragsbepalingen te verklaren. Ondanks dat wij het beide eens zijn dat de regel een organiserende rol kan hebben binnen het sociaal-evolutionaire onderzoek, verklaart het daarom of ondanks dat het oorzaken citeert nog niet. Daarentegen beargumenteren wij de dat regel twee zaken kan bieden: óf een uitleg door beperking óf een niet-causale contrafeitelijke verklaring.