PHILOSOPHY EVOLVING:

SOME PERILS OF EVOLUTIONARY EXPLANATION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Subrena Elaine Smith

January 2014

PHILOSOPHY EVOLVING:

SOME PERILS OF EVOLUTIONARY EXPLANATION

Subrena Elaine Smith, Ph. D.

Cornell University 2014

Many philosophers and psychologists with a naturalistic bent draw on evolutionary biology to support theories about aspects of human psychology and behavior. In this dissertation, I examine and develop some challenges to this approach. The dissertation is intended as an interdisciplinary contribution, examining methodological issues about the study of human behavior and drawing not only on philosophical resources, but also, crucially, on the work of biologists, evolution psychologists, sociobiologists and behavioral ecologists. Chapter 1 addresses Ruth Millikan's theory of proper functions. Millikan uses a model derived from evolutionary biology to give an account of the purposive and quasi-normative features of biological characters. I argue that although the theory of proper functions gives a useful analysis of the purposive and quasi-normative features of biological items, it places undue and unnecessary emphasis on selection at the expense of fitness. I go on to show that Millikan's attempt to reduce intentional purposes to proper functions fails, and

that it conflates ultimate explanations with proximate ones. Chapter 2 develops a novel line of argument against evolutionary psychology, showing that the discipline is riddled with methodological flaws which result from attempts to offer ultimate explanations of contemporary human psychology. I argue that evolutionary psychologists are unable to show that the proximate psychology of contemporary humans is identical to the proximate psychology of prehistoric hominins, because they do not address the problem of individuating modules. I argue that such identities may be impossible to establish, and that any version of evolutionary psychology is therefore likely doomed to failure. Chapter 3 assesses Richard Joyce's evolutionary argument for moral skepticism. I show that evolutionary arguments of the sort that he offers suffer from various scientific and philosophical difficulties, including variations on the problems identified in Chapter Two. Joyce's argument consists of two components: an empirical argument about the evolutionary origin of moral concepts and an epistemic argument purporting to draw a metaethical conclusion from the evolutionary story. I show that both the empirical and epistemic components of Joyce's argument are seriously flawed, and that it cannot help one decide between moral realism and moral anti-realism, as Joyce claims.

BIOGRAPHICAL SKETCH

Subrena E. Smith was born in Westmoreland, Jamaica, obtained her B. A. in philosophy from the University of London (Birkbeck College), and her M. A. in philosophy from Cornell University.

For David Livingstone Smith; you're amazing, stay that way.

ACKNOWLEDGMENTS

I am deeply indebted to Richard Boyd, my advisor, for working with me. Dick's support, encouragement, and belief in me were invaluable to my development. Thank you for your generosity and thoughtfulness. When I entered graduate school I was sure that I would work on metaphysics. That changed when I took Dick's philosophy of science course in my second year. I knew that I wanted to work with him. I wanted to understand how to think about and address methodological questions involving science. He does this better than anyone I've known. Dick's ability to zoom out and see an issue was key to my development.

Thank you greatly Matti Eklund for never giving up, for your mentorship, for your kindness, and most of all, for checking in when I checked out. Your detailed feedback on my work kept me vigilant, made me want to do better, and to be a better philosopher.

Derk Pereboom, you have a way of hearing and asking the right sorts of questions. I appreciate your helping me make clear what you knew I wanted to say. I thank you for being on my committee and for being a valuable teacher.

My parents, Elaine Stewart and Winston Graham, both of whom live in Jamaica, are worthy of special thanks. They toiled greatly on my behalf. They nurtured my curiosity in every way they could. I am grateful for their foresight and love. To my sister Tracian Annmarie Graham, I appreciate all that you've done and continue to do. You are a rock.

To my step-son Benjamin Smith, thank you for the lesson that changed my life's trajectory. I am grateful to Peter Godfrey-Smith for his invaluable guidance when I was

visiting Harvard. Christine Korsgaard, you made me feel like I belonged in the club; thank you. To Gal Kober, I adore you. Thank you for being my friend. It's a pleasure to do philosophy with you. And I must thank Guy Longworth and Samuel Guttenplan. Guy, you are a fine teacher. Sam, thanks for the opportunity.

Thanks also to Paul McNamara and my colleagues in the philosophy department at the University of New Hampshire for their tremendous support.

Zadie Girl Smith, you've been a joy to have in my life. Thank you for your warm love, and for more than you can comprehend.

I arrive at the person whom I owe the most thanks, my beloved David Livingstone Smith. You've been there and seen it all. My learning curve has been steep and it was due in part to you. It was eight years ago that I coyly whispered that I wanted to go to graduate school. I am thankful for the support you gave me. When I was accepted to Cornell you celebrated like it was 2014. You were my rock when after I arrived at Cornell and began to feel the burn of graduate school. I called to say I wanted to come home, you were as ever kind and gentle, but you told me that we had moved and that you would not tell me where we've moved. I am so grateful to you for the hundreds of hours you spent to come to see me. These last years of writing have been difficult, and I was sure that they would be the death of me, but as ever you held me together with your steadfast encouragement; I didn't think so at the time, but thanks for initiating discussions which helped me think and write more clearly. Thanks for keeping the home fires burning with clean clothes, dog care, and my favorite meals. For all that you've done, thank you David. If there is an eternal, I will be eternally thankful to you for

your kind gentleness, your brilliant mind, your ability to keep going, and for your deep abiding love. I love, love you.

TABLE OF CONTENTS

CHAPTER 1

On the Limits of Evolutionary Theory: Millikan on Purpose[1]

1. Introduction: philosophy's evolutionary turn

Incorporating evolutionary theory into one's philosophical enterprise is now in vogue. This is not only the case among philosophers of biology who must at times address questions pertaining to evolutionary biology, but also among philosophers whose work addresses questions outside of biology, narrowly conceived. For many of the latter, the use of evolutionary theory is not a mere ornament (an attractive but dispensable component of their work) but rather is considered as a fundamental theoretical tool (e.g., Gibbard 1990, Skyrms 1996, Dretske 1997; Kornblith 2005, Street 2006, Joyce 2006, Neander 2012).

It is easy to understand the motivation for philosophy's evolutionary turn. We human beings are animals that were shaped by evolution, and this suggests that facts about the kind of biological organism that we are might be useful in addressing questions about our properties and capacities – including questions about traditional philosophical topics such as reference, knowledge, mental content, and how it is that we are able to live purposeful lives. Evolutionary biology is an enterprise that has fundamentally

---

[1] In this chapter, I use the word "purpose" in several ways. In biological contexts, or

altered the way that humans conceive of their position in the natural world. Might it not also have the power to reconfigure how we think about ourselves philosophically? The biological facts must, it seems, have *some* bearing on how we explain *some* things about human beings, but exactly what explanatory role or roles these facts should play in philosophical theorizing remains obscure.

In the present chapter, I will engage with this problem by investigating Ruth Millikan's use of evolutionary biology. I wish to determine how she uses evolutionary theory philosophically, as well as what she gets right and what she gets wrong when she applies it. Her work promises to make for an informative case study, because it is perhaps the most thoroughgoing, scientifically informed, and ambitious attempt to harness evolutionary theory for philosophical purposes.

I will focus on the most central component of her system: her theory of proper functions. The chapter consists of two parts (in addition to this introduction, the conclusion, and a list of references). In the first part (Sections Two and Three) I set out and motivate the theory of proper functions, using the problem of content as a point of entry. I begin with a discussion of the problem of explaining mental content, with an emphasis on the importance of accounting for misrepresentation, and segue into Millikan's claim that one needs an empirically respectable analysis of the normative dimension of biological functions to explain misrepresentation. I then delve into her analysis of natural purposes and explicate three distinct kinds of biological explanations. In the second part of the chapter (Sections Four through Nine) I assess Millikan's project, with a focus on whether the explanations entailed by her theory are reductive. In this part of the

chapter, I focus entirely on the biological adequacy of her theoretical apparatus, on the grounds that any appropriation of evolutionary theory for philosophical purposes will fail unless it gets the science right. I begin with a discussion of what it is for an explanation to be biologically reductive, and identify three arenas in which reduction can occur (corresponding to the three kinds of biological explanations described earlier). Next, I argue that the selectionist component of the theory of proper functions leads to conclusions that are not only deeply unintuitive, but are also reductive, and inconsistent with the correct role of natural selection in biological explanations. I then go on to query whether a Millikan-style theory is reductive in other respects by testing the approach (shorn of its selectionist overlay) against challenges posed by biological learning, novel adaptive behavior, and developmental plasticity. I argue that the theory of proper functions is compatible with our best scientific understanding of these phenomena. Last, I evaluate Millikan's claim that intentional purposes are reducible to biological proper functions. I show that the claim is undermotivated, and that it is inconsistent with her theory of proper functions. Examining this overextension of biological explanation throws light on the limits of evolutionary explanation.

2. Motivating Millikan's project

To understand Millikan's evolutionary turn one must get to grips with her broad philosophical platform. In asking philosophical questions about human beings and their capacities, she never loses sight of the fact that human beings are animals. At a bare minimum, the fact that we are animals places constraints on how we can go about answering certain philosophical questions. Philosophical explanations need to be

consistent with the best current accounts of biological systems (an elegant theory of, say, mental representation that is inconsistent with the way that the mammalian nervous system works is, in effect, no theory at all). But Millikan thinks that the philosophical importance of biology is much more profound and far-reaching than merely providing a constraint on theorizing. She holds that some philosophical questions are not answerable unless we draw on biology to answer them. She does not begin by asking what the relevance of evolution is to some philosophical question. Rather, she starts with some feature of organisms or their relation to the world that is difficult to make sense of. She then asks what must be the case in order for the phenomenon to obtain and, after considering the projectable alternatives, is ultimately led to the conclusion that evolutionary theory (applied either literally or, as we will see, metaphorically) is needed to explain it.

It is helpful to set out Millikan's theoretical apparatus in the context of how she addresses a particular philosophical problem. Consider the problem of explaining mental content. To explain mental content, one needs to identify the sort of relation that obtains between electrochemical states in the brain and states of affairs, such that the former have the latter as their content. The nature of this relation is puzzling, because it seems to be different in kind from all of the other relations that are recognized by science (causal relations, spatial relations, temporal relations, and so on). Millikan addresses this difficulty by identifying biological items which, although non-mental, nonetheless have content. She then develops an analysis that accounts for *their* having content, and applies this analysis to address the problem of mental content.

Millikan argues that it is reasonable to think of intentionality as a property that is shared by many biological systems, rather than something that is restricted to the realm of the mental. Her approach is guided by the broadly Darwinian assumption that all features of organisms are products of evolution, that the differences between organisms are the cumulative result of very many small genetic changes over vast expanses of time, and that comparative methods can sometimes help us understand characteristics that members of different clades have in common. If a characteristic is widespread among organisms belonging to different clades, then this strongly suggests (but does not entail) that they share the characteristic because of their common descent. Biologists use the term *homology* for similarities across clades that are explained by common descent (see paper two of this dissertation). Understanding this, and understanding the ecological pressures that gave rise to the characteristics, can allow one to make inferences about its genealogy and nature.

One of Millikan's methodological innovations is to use cladistic analysis philosophically. Take some puzzling feature of human beings – for example, our capacity to token contentful states. Next, identify what appears to be a non-human homolog of that feature. Then use the homologous trait to make inferences about the human trait. She uses this method to explain mental content by focusing on the signaling behavior of organisms that cannot reasonably be thought of as having propositional attitudes. Her (1984) paradigm case is the waggle dance performed by honeybees (*Apis mellifera*). When a honeybee returns to her hive from a successful foraging expedition, she performs a series of stereotyped movements. These movements systematically co-vary

with the distance and direction of the nectar source, as well as the quantity of nectar to be found there. The dance guides other bees to the source of nectar by providing them with information about its location. This information is encoded in a figure-eight pattern on the vertical comb. The pattern consists of "waggle runs," during which she rapidly shakes her body, alternating with return phases in which she loops back to her starting point. The dance encodes information in several ways. The orientation of the run gives information about the direction of the food source relative to the position of the sun. The duration of the run gives information about the duration of the flight to the food source, and the duration of the whole dance gives information about the quality of the food source.[2] The level of detail that is included in the discussion above should not be taken as extraneous to the goals of the chapter as a whole. Rather, its aim is to steer the reader into Millikan's framework.

It looks like honeybee dances are *about* nectar. But it is not reasonable to think that these insects' nervous system can support propositional attitudes. This, in conjunction with the thought that honeybee signals (or the neural mechanism that produces them) are homologs of mental contents, suggests that whatever it is that accounts for honeybee dances having content might also account for mental states having content.

---

[2] Gould (1975) shows that the systematic, combinatorial properties of honeybee dances are such that they can generate at least forty million distinct signals, and Crist (2004) argues that the waggle dance ought to be regarded as a genuine language, but see Anderson (2004) for objections.

It is intuitive that honeybee dances have the purpose of communicating information about nectar sources to hive-mates. It is this that makes it possible for dances to succeed in achieving their purpose or fail in doing so. That honeybee dances and other non-human signals have purposes in this sense, and therefore are apt for success or failure, suggests that normativity or something like it is a general characteristic of biological systems rather than something that is restricted to humans. If this broad view of normativity is right, it rules out accounts of normativity that are solely concerned with humans. Consider the problem of explaining how humans misrepresent. According to one view, misrepresentations have content, but their content consists of merely intentional states of affairs.[3] But appealing to intentional objects as an explanation of misrepresentation *simpliciter* is inadequate. It is inadequate because it cannot be applied to every species that misrepresents. Suppose that a honeybee executes her dance incorrectly and thereby misrepresents the location of nectar. If bees do not have mental representations,[4] then an incorrectly executed dance cannot represent a merely intentional nectar source.

Millikan argues that making sense of honeybee dances and a whole range of other biological phenomena requires an analysis of their normative dimension. Such an

---

[3] This view is often incorrectly attributed to Brentano. See Crane (2006).

[4] In what follows, I will refer to any contentful items as "representations." Millikan initially limited "representation" to cover states of organisms that are about the world and which participate in inferences and re-identifications, and distinguished them from simpler "intentional icons." In her later writings, she adopted the convention of referring to both of these as "representations".

analysis needs to clarify what it is about them that accounts for their being the sorts of things that can succeed or fail. She addresses this by means of her theory of proper functions.

3. Proper functions

Unlike the objects investigated in physics and chemistry, biological items seem to have purposes (in a sense that I shall shortly make clear). Honeybee dances do not just signal the location of nectar; they are *for* signaling the location of nectar. Hearts do not just pump blood; they are *for* pumping blood. And eyes do not merely see; they are *for* seeing. Such attributions of purpose come very naturally. But what should one make of them? Are these items really purposive, and are there purposes in nature more generally? If there are, how can one account for them in a manner that is consistent with the scientific world-view?

Saying of something that it has a purpose is ambiguous. On one reading, having a purpose is the same as having an intention. For example, one has the purpose of going outside if one's behavior is guided by the intention to go outside. Clearly, *this* notion of having a purpose can only be applied to whole organisms that possess the sort of cognitive architecture that makes it possible for them to have intentions. It would be laughable to say that bee dances have purposes in this sense.

The purposes that one attributes to honeybee dances and bodily organs resemble the purposes attributed to artifacts more closely than they do the purposes attributed to agents. When one says of a clock that it has the purpose of telling time, one does not mean to say that the clock intends to tell time. Rather, one means that it is designed to

tell time. This is why it makes sense to evaluate the performance of clocks (*qua* clocks) on the basis of their performance as time-keepers. But there is an important sense in which the purposes of artifacts like clocks are dissimilar to the purposes of biological items. The purposes of artifacts are derived from their designers' intentions. But biological items are not intentionally designed. They come about through the purposeless process of evolution. Suppose, then, that one treats evolution as a grand designer, and says of biological items that what they are for is what evolution designed them to do? Suppose that honeybee dances are for signaling the location of nectar because Mother Nature "designed" them for this purpose. The metaphor of nature as an engineer is evocative, and has gained wide currency (e.g. Dennett 1995) but it is not informative unless the "designer" metaphor can be cashed out empirically.

Millikan's key biological notion — the foundation of her whole philosophical apparatus — is her theory of proper functions. To have a proper function, an item must belong to what she calls a *reproductively established family.* These are populations that are united by common descent. Clades are reproductively established families, and so are anatomical categories like "mammalian heart" and behavioral categories like "honeybee dance." *First-order reproductively established families* are created by the reproduction of a prototype, and reproduction of reproductions of the prototype. A population consisting of tokens of an allele is a first-order reproductively established family if every one of those alleles is either a copy of the original allele or a copy of a copy of it. *Higher-order reproductively established families* are populations of items that are produced by first-order reproductively established families. So, my left foot is a member

of a higher order reproductively established family. It is not a copy of my mother's left foot or my father's left foot, but it was built by genes that were copied from my parents' genes. Left feet do not have ancestors in the ordinary sense of the word. However, they have something like ancestors in an indirect, derivative sense – a sense in which my grandmother's left foot was an ancestor of my left foot. In this chapter I will use the terms "ancestor" and "ancestral" as Millikan does to cover straightforward first-order cases like genes as well as higher-order cases like left feet. It is important to note that this departs from the orthodox use of these terms.

There are two kinds of proper functions: direct ones and derived ones. The direct proper function of a trait is that for which it was selected for doing in previous generations. These traits were selected for because their effects accounted for their own reproduction in an ancestral population. Some items with direct proper functions were selected for performing their functions by means of producing other items. Items that are so produced derive their proper functions from the items that were selected for producing them. Consider the human pancreas. One of the direct proper functions of the pancreas is to regulate levels of blood sugar. It realizes its proper function by producing the hormones glucagon and insulin, which modulate levels of blood sugar (glucagon raises it and insulin lowers it). Pancreatic hormones derive their proper function of regulating blood sugar from the proper function of the pancreas because, in producing them, the pancreas performs its proper functions. It is important to note that selection plays an essential role with respect to both kinds of proper function. By Millikan's lights, all proper functions have to be selected for.

4. Considering an alternative

Here is where things stand. Millikan's treatment of the problem of content places biology at the center of her solution. I think that her strategy has some obvious benefits, but it also has some apparent disadvantages. There is a view amongst some philosophers of science that scientific explanations must be causal explanations (e.g., Salmon 1984, Lewis 1986). On this view, biological norms can play no role in scientific explanation because they are causally inert. But even if one countenances the scientifically legitimacy of some non-causal properties, framing ones explanations in terms of straightforwardly causal properties may be preferable on methodological grounds. Consequently, it is important to consider whether there is any alternative to Millikan's explanations of natural purposes that does not rely on non-causal properties. Broadly speaking, there are two such alternatives. One way to proceed is to reduce biological norms to ordinary causal phenomena. The other is to eliminate them entirely.

Both reductionists and eliminativists draw on a causal-role account of biological functions. This account, as developed by Cummins (1975), analyzes the functions of biological items as their causal contribution to capacities of larger biological systems of which they are parts (hence, it is often referred to as the theory of systemic functions). A number of philosophers[5] endorse both the theory of systemic functions and the theory of proper functions, because they regard them as having distinct and mutually

---

[5] I do not wish to imply that Millikan's and Cummins' theories are the only accounts of function currently on the table. See, for example, Neander (1991a, 1991b), Papineau (eliminativists 1984), Dretske (1986), Schroeder (2004), and Mossio et. al.(2009).

compatible explanatory roles. But reductionists hold that systemic functions can do all the work that proper functions do, including accounting for biological normativity, and they argue that this approach is preferable on grounds of parsimony. Eliminativists accept that the systemic theory cannot account for biological norms, but argue that this is not a problem for it because biological norms are illusory.

I begin with eliminativism. Pargetter and Bigelow (1987) express the eliminativist worry as follows, "It is assumed that functions, if there were any, would have to be important, *currently existing*, causally active and explanatory properties of a character or structure….It is thus concluded that there really are no functions in nature" (182-183).

> To add functions to the scientific biological picture, on this view, is parallel to adding final causes to physics. Final causes have no place in the scientific account of the physical universe, and, if the psychological pressures are resisted, we find we can do without them and final causes just fade away. To the eliminativist, the same will be true of functions; as the biological sciences develop, any need for function talk will vanish, and the psychological naturalness of such talk will fade away with time and practice (183).

This project can only succeed if, as Davies (2003) argues, biological science can dispense with the notion of proper functions at no explanatory cost, because there are no facts about organisms that would be inexplicable without it. For this project to go through, it must be the case that *objectively speaking* biological items cannot succeed or fail (in the normative sense). Of course, one might say that they succeed or fail to do what one expects them to do or want them to do, but these would be subjective rather

than objective normative criteria. Imposing normative standards on natural phenomena is a different matter from finding them there. There are no norms in nature.

Contrary to the claim that norms do no explanatory work in biology, it is clear that eliminating them would impoverish biological discourse, because unlike explanations in physics and chemistry, explanations in biology make essential use of notions such as *damage*, *malfunction*, *illness*, and *deformity* – notions that are intelligible only with reference to standards of performance. These notions are not extraneous to the inferential practices that are characteristic of biology. They are biologically explanatory, and their being so presupposes realism about such norms, without this entailing any particular view of what they consist in.

For eliminativists, getting rid of biological normativity entails getting rid of notions like malfunction. For Boorse (2002),

> If Carla's heart cannot pump blood, then pumping blood is not, in fact, the function of her heart; it has no function. Since pumping blood is the [statistically] normal function of a human heart, it would be the function of Carla's heart if Carla's heart functioned normally; but it does not, so it is not (89).

It is standard practice among medical professionals to make claims of the form "Some feature of x is not functioning as it should." They make such claims because they have already bought into the idea that there are normal and abnormal ways for the item under consideration to behave. It is precisely because pumping blood is its function that one holds that it is a problem that Clara's heart has stopped pumping blood. Judgments of this kind motivate the development and implementation of procedures for restoring

normal function. With notions like “malfunction” off the table, it is not clear how medical scientists could carry out their work. The eliminativist stance is extreme, and, if adopted, would cause more problems than it would solve. Reductionism might be a more promising strategy.

Godfrey-Smith (1993) argues that malfunctions can be understood in purely causal terms. They can be analyzed as departures from what is typical of their kind. If a biological token is not able to fill the causal role that tokens of their kind usually fill in broader systems of which they are parts, then the token malfunctions. If this analysis is sustainable, then the theory of systemic functions provides a live alternative to the theory of proper functions. There are several problems that any version of the typicality theory must surmount. Any such theory must have the resources to explain how it is that there are biological items that typically do not perform their functions. It is uncontroversial that sperm cells are for fertilizing ova, even though fertilizing ova is far from typical of sperm cells.[6] This shows that typicality is not *necessary* for biological norms. Additionally, the criterion does not distinguish functions from their side-effects. Making thumping sounds is typical of hearts, but failing to make a thumping sound is not a malfunction (it is an *effect* of a malfunction). A third worry is that the typicality analysis does not distinguish failure from abnormal success. Any performance might deviate from the (statistical) norm either because it underperforms or because it performs exceptionally well (see Amundson 2000). Both better-than-average vision and worse-

---

[6] The example is from Millikan (1984).

than-average vision are deviations from the statistical norm. By the typicality criterion, both must be characterized as malfunctions. A fourth difficulty concerns the problem of individuating populations. Traits count as typical or atypical only with respect to populations. So statements about typicality are only contentful if they are statements about typicality-in-a-population. This raises an important question about how to individuate populations. Should the reference populations for typicality judgments be whole species, organisms in certain ecologies, organisms at a certain stage of their life cycle, or some other division? An item might typically behave one way in population A and another way in population B, even though B is a subset of A. In that case, there are two different, and possibly inconsistent, norms of functioning for a single item. For example, the rate of myopia worldwide is around thirty percent, but in Singapore it is as high as eighty percent (Seet et al. 2001). So, myopia is not typical of people generally (at least, at the present time), but it is typical of the residents of Singapore. So, by the typicality criterion, myopia is a disorder when considered in relation to the world population as a whole and not a disorder when considered in relation to the population of Singapore. Consequently one cannot say, without qualification, of any nearsighted Singaporean that her eyes malfunction.

In conclusion, eliminativism about biological norms does not seem plausible, and the strategy of reducing norms to typicality is faced with serious obstacles. As these appear to be the two most plausible alternatives, a strategy like Millikan's, which grants that purposiveness is an irreducible feature of certain biological phenomena, is *prima facie* vindicated. It is important to note, though, that this conclusion endorses the general

form of her solution, not the details of her analysis (about which I will raise objections) or her manner of applying it.

5. Three kinds of biological explanation

To properly assess the role of biology in Millikan's system, and to evaluate the extent to which her philosophical strategy succeeds, it is necessary to examine her notion of biological explanation more closely.

The theory of proper functions abstracts away from key elements of the theory of evolution. In abstracting away from the biological details, Millikan insists that her theory goes further than biology in the literal sense does. The process of evolution is, she believes, just one instantiation of the broader proper function formula, which is also realized in other domains, such as psychology and culture.[7]

Somewhat confusingly, Millikan tends to characterize *anything* that satisfies the proper function formula as "biological." Her indiscriminate use of the term sometimes makes it difficult to know exactly what claim she means to be making. Consider her (1984, 2005) claim that language is "biological". There is a trivial sense in which language is biological. As far as is known, *Homo sapiens* are the only language users. They are animals, and animals are biological entities, so (platitudinously) language is a biological phenomenon – but then so is every feature of every organism. But Millikan does not have this vacuous notion in mind when she describes things as "biological," nor does

---

[7] The idea the evolution can be abstractly characterized comes from Lewontin (1970). See Buskes (2013) for a useful survey and evaluation of views on evolution as a multiply realizable, "substrate-neutral" process.

she mean to say that all "biological" items are literally biological either. For example, she writes in *Language, Thought and Other Biological Categories* that she "used 'biological categories' by extension to cover all proper function categories" (1984: 29). And in her 1993 essay "Propensities, exaptations, and the brain," she informs readers that the point of the definition of proper function was to "capture a certain similarity that I took to be important among items falling in [real] biological categories, language categories, purposive action categories, artifact categories, and certain kinds of cultural categories" (1993b: 31). So, "biological categories" in a narrow sense – categories like "heart" or "mitochondrion" – are said to fall under "biological categories" in the broad sense only insofar as they possess proper functions. She then adds:

> It was probably unwise to use the term 'biological category' as an informal substitute for 'proper function category'...for this has given the false impression that my aim was to capture biologists' usage. On the contrary, I do not consider biological examples of proper functions to be more central than any others (1993b: 31).[8]

Sifting through the various examples scattered through her writings, it becomes clear that Millikan offers three distinct sorts of biological explanations. All of them cite evolution, or evolution-like processes, but they do so in different ways. Some explanations are what I call *strictly* biological. These are paradigmatically biological

---

[8] Elsewhere, she states explicitly that the term "biological" as used in the title of *Language, Thought and Other Biological Categories* "is used not literally, but broadly or metaphorically" (2002: 115).

explanations of paradigmatically biological phenomena. She also occasionally offers what I call *extended* biological explanations – paradigmatically biological explanations of ostensibly non-biological phenomena (a kind of explanation that is commonplace in the literatures of evolutionary psychology and human sociobiology, as well as the work of philosophers who are influenced by these disciplines [e.g., Street 2006, Joyce 2006]). Both strict and extended forms are *literally* biological, but Millikan also gives what I call *metaphorically* biological explanations, which attribute proper functions to items that do not fall within the scope of literally biological explanations. Consider her (2006) account of operant conditioning. When a rat is rewarded by a food pellet whenever it pushes the bar in a Skinner box, it learns to push the bar repeatedly. It is natural to say that once the bar-pushing becomes established in response to the reward that it produces, the bar-pushing acquires the purpose of getting food. However, in saying this one does not have to assume that the rat intends to get food by pushing the bar. An explanation of the rat's behavior that is cast in terms of proper functions enables one to attribute a purpose to the behavior without defaulting to an intentionalistic explanation of it: the behavior has a purpose even though it is not the rat's purpose to perform the behavior. The rat's behavior is said to satisfy the proper function formula because (a) the rat pressed the bar as well as performing other behaviors, (b) the first bar-pressing was reproduced because of its effect, and (c) bar pressings reproduced more successfully than non-bar-pressings. Like all metaphorically biological explanations in Millikan's work, the proper function of the rat's behavior is cast in an evolutionary biological *form* but it does not have evolutionary biological *content*.

Given the role of selection in the proper function formula, each of these kinds of explanation requires that the item being explained was selected for, either literally or metaphorically. When Millikan offers strict and extended biological explanations, she assumes that the item being explained was subject to natural selection. When she offers metaphorically biological explanations, she has it that the item being explained was subject to a process that bears an abstract resemblance to natural selection – some form of selection that is realized in a psychological or cultural medium.[9]

6. Reduction and the role of selection

One question that can be asked of any biological explanation is whether it is reductive. Accordingly, I wish to explore the question of whether and in what respects Millikan's biological explanations are reductive. There are many kinds of biologically reductive explanations (Brigant & Love 2007). In this chapter, I will consider an explanation to be reductive if it purports to explain some phenomenon exhaustively or near-exhaustively in a more restrictive and simplifying manner than other well-confirmed explanations do. It follows that explanations are never reductive all on their own; they are only reductive in relation to other explanations. Explanations that are reductive in this sense are

---

[9] The idea that there is such a thing as cultural evolution – roughly, that a broadly Darwinian model can explain phenomena in the cultural domain – is not unique to Millikan's work. The idea that an evolution-like process of variation and selection can explain certain kinds of cultural, scientific, and technological change has been advanced by a number of thinkers, including Popper (1972), Toulmin (1972), Campbell (1974a, 1974b) Wilson 1975, Dawkins (1976), Cavalli-Sforza & Feldman (1981), Boyd & Richerson (1985), Hull (1988) and Dennett (1995). See Buskes (2013) for a review and evaluation of this theoretical tradition.

scientifically dubious, because they fail to address the evidential and inferential considerations that motivate projectable rival accounts.

All three kinds of biological explanations that I discussed above can be applied reductively. I will consider each in turn. Strictly biological explanations are reductive if the reducing explanation and the reduced explanation are both paradigmatically biological explanations. Extended biological explanations are by their very nature reductive, because they are literally biological explanations of ostensibly non-biological phenomena. Metaphorically biological explanations are reductive if they purport to account for non-biological phenomena more fully than rival non-biological explanations do. It is an open question whether, in any given case, a reductive explanation succeeds. The warrant for a reductive explanation can be determined only by comparing the explanatory power of the reduced explanation with that of the reducing one.

I wish to argue that the theory of proper functions is reductive in virtue of placing undue emphasis on selection, and that this seriously compromises the theory's credibility. I will also argue that the theory of proper functions would be improved by removing its selectionist element. However, given Millikan's insistence on the indispensability of selection, incorporating such a modification would amount to proposing an alternative theory of proper functions.

Recall that Millikan claims that anything with a proper function *must* have been selected for performing that function. For example, she asserts that "Only if an item or trait has been *selected* for reproduction, *as over against other traits*, *because* it sometimes has a

certain effect, does that effect count as a [proper] function" (1993b: 35, emphasis in original). However, she sometimes does not mention selection at all when describing proper functions. Consider her claim that the proper function of a characteristic is "a function that its ancestors performed that has helped account for the proliferation of the genes responsible for it, hence helped account for its own existence" (1993c: 14). In this description it is fitness rather than selection that is doing the explanatory work. What looks like an oscillation between two alternative formulations is probably more apparent than real. I think that it is best explained by the fact that anything that is selected for *must* promote fitness, because selection is just differential fitness at the level of populations. I suspect that Millikan so takes for granted the intimate tie between fitness and selection that there is an unarticulated assumption that whenever fitness is maximized selection comes along for the ride. But that this is by no means necessarily the case is made evident by the following example. Suppose that there is a species of *drosophila* living in an environment that periodically gets very cold. Some individuals in the population have thick skin, which protects them from the cold, while others have thin skin, which offers no such protection. Suppose also that the individuals with thin skin are just as fertile as those with thick skin when exposed to normal temperatures, but become dormant and cannot reproduce when subjected to very low temperatures. Under these conditions, thick skin is selected for. That is, thick-skinned flies produce more descendants than their thin-skinned sistren, and their numbers increase more rapidly than do the numbers of the thin-skinned flies. After a number of generations,

there are no thin-skinned flies left. The allele responsible for thick skin has reached fixation.

If this were the end of the story, it would be a straightforward Millikanian case. One would say that because the thermal effect of thick skin caused it to be reproduced more effectively than thin skin, and because thick skin was selected for its thermal effect, that thick skin had the proper function of protecting the flies from cold.

Now, suppose that the following occurred. At some point after the gene causing thick skin reached fixation, there was a change in the climate. Very low temperatures no longer occurred, and the warmer conditions caused deadly parasites to proliferate – parasites that preyed on insects. The parasites normally enter their hosts' bodies by boring through their skin, and once inside, they multiply and kill their host. However, thanks to the flies' thick skin the parasites were unable to infect them.[10] Under these altered environmental circumstances, thick skin provided an important fitness benefit, and this benefit undoubtedly explained its continued reproduction, but this was not the fitness benefit for which the trait was selected. According to Millikan's theory, the proper function of thick skin remained that of protecting flies from the cold, because that is what it was selected for doing. But this diagnosis seems wrong. It is more intuitive that its proper function changed to protection from parasites, even though there was no

[10] This example is a hybrid of a case described by Dover (2000) and a case described by Sober (1984).

selection for protection from parasites.[11] It is clear that if they had not had thick skin the flies would not have been able to reproduce at all, since the parasites would have driven them to extinction. In fact, in these circumstances, thick skin offered greater fitness benefits than it did in the circumstances where it was selected for. In the first situation, thin-skinned flies could reproduce, but could not reproduce as effectively as thick-skinned flies. But in the second situation, thin-skinned flies would not have been able to reproduce at all.

Millikan does not offer any justification for the selection requirement. The core of the proper function formula – the idea that the purpose of a trait can be explained by its contribution to its own reproduction – is all about *fitness*. And the selection clause does not to add anything useful to the analysis. To see why, one need only consider what natural selection is. Natural selection is a way of describing effects of phenotypic traits at the population level. Saying that a trait has been selected for performing some function is equivalent saying that the trait is more fully represented in the population than some alternative trait on account of the aggregate effects of the two traits on individual fitnesses. But the representation of a trait in a population does not tell us anything about individual organisms. Okrent (2007) puts the point admirably well.

> Population thinking accounts for the prevalence of some structure or behavior in some population, not for the presence of some structure or behavior in some individual, as Millikan thinks. One accounts for the fact that a certain percentage

[11] For related discussions see also Buller (1998), Abrams (2006), and Preston (2009).

> of humans have a heart of a certain type by appealing to the facts that some ancestor population had a range of different hearts and that, in the environment in which those ancestors functioned, those with the kind of heart that most of us now possess were more fit than the others….But what is explained is *always* that some percentage of some population, perhaps approaching 100 percent, has such and such features (95).

Selection cannot enter into an explanation of phenomena at the level of individuals. It cannot explain, and should not be used to explain, why it is that a particular trait has a certain proper function in the lives of organisms that possess the trait. By analogy, knowing that in a population of coin tosses heads come up roughly fifty percent of the time does not explain why any particular toss comes up heads. Millikan's selectionist account of proper functions is reductive, because it misconstrues the explanatory role of natural selection by conflating population-level explanations with explanations at the level of individuals (individual organisms, structures, behaviors, or genes). Millikan neglects to show that fitness cannot do the job that she tries to recruit selection to do. Orthodox Millikanian explanations are reductive even in cases where the proper function of a trait is what the trait has been selected for doing, because even in these cases selection is supposed to be constitutive of the trait's proper function.[12]

---

[12] It is not clear whether the selectionist emphasis renders metaphorically biological explanations reductive as well, and I will not explore this question in the present paper

I have argued that Millikan's undue emphasis on selection reveals her conflation of explanation at the level of populations with explanation at the level of causes operating within populations. It may be that this conflation is a side effect of her warranted commitment to the importance of accounting for biological purposes in terms of ultimate explanations. It may be that she assumes that ultimate explanations have to be selectionist explanations (see paper two of this dissertation). This view is shared by many biologists, beginning with Mayr (1988) who is acknowledged as the father of the proximate/ultimate distinction, even though it is unwarranted (Dewsbury 1999, Ariew 2003, Laland et. al 2011). These commitments may have led her to assume that because proper functions require ultimate explanation, they must be framed in terms of selection. However, although it is true that selectionist explanations are ultimate ones, it is not true that if an explanation is ultimate, then it must be framed in terms of selection. Millikan could, if she wished, abandon the selection requirement while offering ultimate explanations that are framed in terms of fitness.

7. Three case studies

I will now consider whether the theory of proper functions is reductive in other respects. I will do this by investigating whether the theory can handle several sorts of adaptive biological phenomena that might be expected to put pressure on it. To avoid complicating matters unnecessarily, I will assume that the proper function of the trait is fixed by its effects on fitness rather than by selection.

I begin with the challenge of explaining how biological learning can produce purposeful behaviors. Some forms of learning are paradigmatically biological (for example,

learning to identify kin by means of imprinting mechanisms), while others are not (for example, learning trigonometry). The difference between them may be explained by the domain-specificity of biological learning and the domain-generality of non-biological learning.[13] The idea is that biological learning is the result of an evolved neural device that is geared toward causing the animal to learn one specific kind of thing, in contrast to learning mechanisms that can be applied more generally to a wide range of learning tasks.[14]

Animal song is an interesting arena for exploring the interface between learning and evolution. Songs are species-specific acoustical signals that are typically produced by males in the context of courtship or territorial defense. Most animals that produce song do not have to learn their songs, but humans, cetaceans, certain bats, and some birds are exceptions (Kroodsma & Miller 1996).

Vocalization by songbirds is a well understood example of learned signaling. Birdsong has a biological purpose; namely, attracting mates and establishing territorial boundaries. Because these purposes are mediated by learning, the theory of proper

---

[13] By this criterion the example of operant conditioning given earlier is not an example of biological learning, because the rat's learning to press the bar is the upshot of a domain-general learning mechanism.

[14] In some cases, the production of signaling behavior is unlearned but appropriate responses to the signal must be learned in whole or in part. Honeybee dances are an example. Although there is no evidence that the dances are learned, learning influences the manner in which observer bees respond to dances (Biesmeijer & Seeley 2005, Grüter et al, 2008).

functions cannot accommodate them unless it is consistent with the existence of learned biological purposes.

Comparative neurobiological research suggests that songbirds are born with an evolved learning device (the "song system"). The song system enables them to produce songs, learn songs, and evaluate and correct the fidelity of the songs that they have learned. It is transmitted genetically, and probably became established in the songbird lineage because singing species-appropriate songs enhanced the fitness of ancestral songbirds (Brainard & Doupe 2002). Given these facts, one can say that the song system has the proper function of causing birds to produce species-appropriate songs. Now, saying that a device has a proper function does not entail anything about the proximate mechanisms by means of which its proper function is realized.[15] These might include a whole range of causal processes, including learning. In the case at hand, the song system fulfills its function by biasing song learning towards species-appropriate songs (Dooling & Searcy 1980, Nelson & Marler 1993, Marler & Peters 1988). In other cases of vocal signaling, the proper function of the neural device is realized by means of different proximate mechanisms. For example, the suboscine passerines, a group of birds that are closely related to the songbirds, do not learn their songs. Their singing patterns are highly canalized and not dependent upon acoustic inputs (Beecher &

---

[15] In Millikanian jargon, an item's having a proper function is distinguished from the historically normal explanation of how items of that type perform the function. An item is said to perform its proper function in a historically normal way when it does so by means of the same proximate mechanisms as were operative in its Environment of Evolutionary Adaptedness.

Brenowitz 2005). Their song system realizes its proper function by means of proximate mechanisms that are different from those in the songbird case, even though in both cases the song system has the same proper function. So, correctly understood, the theory of proper functions does not have unacceptable nativist entailments, and it is consistent with purposeful behaviors resulting from biological learning.

The second topic that I wish to address in this section is the question of whether the theory of proper functions is compatible with the existence of novel adaptive behaviors. These are behaviors that have the purpose of adapting an organism to a current environmental circumstance that had no evolutionary precedent. Consider again honeybee dances. Once the signaling system became established in the honeybee lineage, there must have been many occasions when foraging bees found nectar in locations (relative to the hive) where it had never been previously found. It is reasonable to suppose that bees were able to signal information about these novel locations because they were equipped with a neural system that *systematically* maps nectar locations onto dances. The dance-producing mechanism is an example of a neural device that has the proper function of responding flexibly to a range of possible environmental contingencies. The theory of proper functions has no difficulty accommodating novel adaptive behavior in cases where it is a function of an evolved mechanism that systematically adjusts behaviors to environmental contingencies.

The last challenge that I wish to discuss concerns adaptive developmental plasticity. Developmental plasticity is defined as:

> [A] single genotype's ability to alter its developmental processes and phenotypic outcomes in response to different environmental conditions. Such environmental effects on trait expression can range from modest adjustments to growth rate or tissue allocation in response to resource levels, to dramatic polyphenic switches by which a single genotype can give rise to discrete and often radically different alternative phenotypes (Moczek et. al. 2011: 1).

This might be thought to present a problem for the theory of proper functions. If phenotypes can vary as a function of environmental factors, and these variations are purposeful, then the purposes of phenotypic traits do not seem to have been fixed by evolution. Consider the developmental trajectory of water fleas (*Daphnia pulex*) that have been exposed to kairomones (chemical traces indicating the presence of predators) Daphnia that have been exposed to kairomones from the tadpole shrimp (*Triops cancriformis*) during their development respond by growing helmets, tail-spines, and neck teeth. These features have the purpose of protecting them from predation (Ebert 2011).

The developmental pathways underpinning adaptive plasticity are not well understood. One hypothesis cites mechanisms that buffer phenotypes against environmentally induced developmental disturbances. These mechanisms prevent mutations from being expressed, and therefore may allow mutations to accumulate in a lineage over time. At some point, an environmental change, or a mutation that makes the organism more sensitive to some feature of the environment, might trigger their expression, and this might have dramatic phenotypic consequences. In the very occasional cases where the

resulting phenotype enhances fitness, if the condition that triggers their expression occurs often enough, the mutations responsible for the adaptive trait may persist in the lineage (Moczek et. al. 2011). The trait then acquires a proper function. So, if developmentally plastic traits are preserved in the lineage because they enhanced ancestral fitness, and the manner in which they enhanced fitness is also preserved, then the theory of proper functions has no difficulty accounting for their purposefulness.

8. An overextended biological explanation

The fact that a theory does not *entail* reductive consequences does not prevent its advocates misusing it to draw reductive conclusions. In this section, I discuss one such example taken from Millikan's work. This example is informative because it reveals a reductive component of Millikan's philosophical agenda. But it is even more important because it shows how an example of an overextended evolutionary explanation casts light on the proper limits of such explanations.

Earlier I pointed out that talk about purpose can be ambiguous. On one hand, "having a purpose" can refer to intentional states or intentionally designed items, and on the other, it can refer to non-intentional items or their products. I have also urged that it is plausible that neural systems with non-intentional purposes provide a biological platform for purposes in the first, intentional sense. It seems reasonable to suppose that our capacity to entertain a belief-desire psychology grew out of neurological structures that contributed to our ancestors' fitness. This modest claim does not entail that the underlying neurology was selected for, or even that it enhanced our ancestors' fitness by making it possible to form intentions.

But one might wonder whether the ambiguity noted above points to a deeper relationship between the two kinds of purposes. Perhaps proper function theory can do more than merely explain non-intentional aims and their associated norms. Perhaps it can give us an account of purposiveness and normativity *per se* – including intentional purposes and prescriptive norms. Millikan thinks so. Her aspirations for the theory of proper functions go well beyond using it to account for how non-intentional items can be purposive and subject to norms. She believes that the theory can yield an account of intentional purposes and their related norms. If this project can succeed, it would be a hugely significant philosophical achievement. But can it succeed? I will argue that it cannot, and that attempts to extend the theory of proper functions to the intentional sphere overstep the limits of biological explanation.

Starkly put, Millikan wishes to argue that *all purposes are biological purposes*. "My thesis," she writes, "will be that the unexpressed purposes that lie behind acts of explicit purposing [i.e., intentional purposing] are biological purposes…. Biological purposes are, roughly, functions fulfilled in accordance with evolutionary design" (1993a: 217) and "the normative element that is involved when one means to follow a rule is biological purposiveness" (222).

In her most explicit discussion of the relation between intentional purposes and biological ones, Millikan (2004) contrasts the biological purpose of the eye-blink reflex with the intentional purpose of allowing an ophthalmologist to administer eye-drops. She states, plausibly enough, that the purpose of the eye-blink reflex is to prevent foreign objects from entering the eye. She then goes on to say that our irresistible

tendency to blink even though we desire the medicine to enter our eyes is best understood as a case of *conflicting purposes*, and that both of these purposes are the same kind of thing.

> Maybe you will object that only one of these crossing purposes is a *real* purpose. The other is a "purpose" not literally but only by analogy or metaphorically. The real purpose is the conscious human intention not to blink. Only the intention not to blink is a purpose of the whole person, rather than merely a "subpersonal" purpose. The purpose of the eye-blink reflex is only a "subpersonal" or "biological" purpose, and these are purposes only metaphorically (3).

Given that the purpose of the eye-blink reflex is just its proper function, the claim that proper functions are *real* purposes is the claim that they are metaphysically on a par with intentional purposes. She goes on to claim that the proper functions of parts of humans and the intentional purposes of whole humans, although seemingly distinct, are actually two manifestations of the very same thing. "[N]o interesting theoretical line can be drawn between these two kinds of purposes," she writes, *because* "purposes of the whole person are made up out of intertwined purposes at 'lower' or more 'biological' levels" (ibid.).

There are two reasons why this picture should be rejected. The first concerns her claim that proper functions constitute intentional purposes, and the second concerns her claim that there is no interesting theoretical difference between biological purposes and intentional purposes because the latter are constituted by the former.

With regard to the first objection, notice that intentional purposes are mental *states* with causal powers. For example, having the intention to make a pot of coffee is being in a state that disposes one to make a pot of coffee. In contrast, biological purposes are non-causal, historically determined standards of functioning. Millikan's explanatory framework is predicated on the idea that (proximate) states come apart from (ultimate) biological purposes. That it is possible for biological items to fail to do what they are for doing is what accounts for the normative dimension of biological systems. It light of this it is difficult to make sense of Millikan's claim that there is no fundamental theoretical difference between intentional purposes (causally efficacious states of mind) and biological purposes (noncausal standards of performance). Her position amounts to an abandonment of the very theoretical apparatus that supposedly underpins it.

With regard to the second objection, it cannot be the case, as Millikan claims, that an item's being composed of biologically purposive parts is sufficient for its being intentionally purposive. Non-intentional systems (for example, kidneys) are also composed from items with proper functions, but kidneys are not intentionally purposive. This is a special case of the much more general principle that the properties of a thing cannot be inferred from the properties of its parts (of which Millikan is well aware). So she cannot intelligibly claim that mental states are intentionally purposive *because* they are constructed out of biologically purposive parts.

In conclusion, Millikan's effort to reduce intentional purposes to proper functions cannot succeed. In attempting to do this, she not only oversteps the limits of evolutionary biological explanation, but also undermines the foundations of her own theoretical

apparatus. In her hands, biological explanation can, at least in principle, take one to the threshold of intentional psychology by giving an account of the biologically purposive neural systems that provide the basis for intentional projects and instrumental norms, but it does not have the resources to explain the *content* of those purposes and norms by citing their evolutionary history.

9. Conclusion

My goal in this paper has been to evaluate the strengths and limitations of Ruth Millikan's philosophical appropriation of evolutionary biology. I have shown that her key use of evolutionary theory is to give a naturalistic account of non-intentional norms and purposes both inside and outside the biological domain. This is an important achievement. However, in insisting that items can have proper functions only on the condition that they have been selected for, Millikan conflates a description of the proliferation of traits in a population with an account that explains the fitness-enhancing characteristics of biological characters. This selectionist aspect of her theory is both reductive and also leads to unintuitive conclusions about the purposes of features of organisms. This shortcoming aside, the theory of proper functions emerges as a powerful and flexible tool that is consistent with a scientific understanding of a range of biological phenomena, including domain-specific learning, developmental plasticity, and adaptively novel behavior. However, as is shown by Millikan's attempt to give a reductive analysis of intentional purposes, the theory comes apart when it is pushed beyond its proper explanatory limits (as all theories do).

The results of my analysis of Millikan's project can be distilled into two very general lessons about the philosophical use of evolutionary theory. The first is that, suitably amended, the theory of proper functions provides a robust, empirically defensible strategy for addressing a wide range of purpose-like and normative-like phenomena in several arenas. The second is that her failed attempt to undo the distinction between personal and subpersonal levels of analysis, and to give a reductive account of the former in terms of the latter, underscores the importance of resisting the temptation to use evolutionary theory to address questions that can only be addressed by citing proximate mechanisms (by conflating ultimate why-explanations with proximate how-explanations).[16] Unless one keeps one's eyes firmly fixed on this methodological ball, one ends up trying to force evolutionary theory to do more explanatory work than it can possibly accomplish.

---

[16] This kind of category error is rampant in the literature of evolutionary psychology as well as philosophical work predicated on the "results" of evolutionary psychological research, and even very sophisticated thinkers like Millikan sometimes fall victim to it (See Buller 1999).

REFERENCES

Abrams, M. (2005). Teleosemantics without selection. *Biology and Philosophy*, 20: 97-116.

Amundson, R. (2000). Against normal function. *Studies of History and Philosophy of Science, Part C,* 31: 33-53.

Anderson, S. R. (2004). *Doctor Dolittle's Delusion: Animals and the Uniqueness of Human Language*. New Haven, CT.: Yale University Press.

Ariew, A. (2003). Ernst Mayr's ultimate/proximate distinction reconstructed and reconsidered. *Biology and Philosophy*, 18: 553-565.

Beecher, M. D. and Brenowitz, E. A. (2005). Functional aspects of song learning in songbirds. *Trends in Ecology and Evolution*, 20: 143-149.

Beismeijer, J. & Seeley, T. (2005). The use of waggle dance information by honeybees throughout their foraging careers. *Behavioral Ecology and Sociobiology*, 59: 133-142.

Boorse, C. (1976). Wright on functions. *The Philosophical Review*, 85: 70–86.

Boorse, C. (2002). A rebuttal on functions. In Ariew, A., Cummins, R. & Perlman, M. (eds.), *Functions: New Essays in the Philosophy of Psychology and Biology*. New York: Oxford University Press.

Boyd, R. & Richerson, P. J. (1985). *Culture and the Evolutionary Process.* Chicago: University of Chicago Press.

Brigandt, I. & Love, A. (2012). Reductionism in Biology. In Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/archives/sum2012/entries/reduction-biology/>.

Brainard, M. S. & Doupe, A. J. (2002). What songbirds teach us about learning. *Nature*, 417, 351-358.

Buller, D. J. (1998). Etiological theories of function: a geographical survey. *Biology and Philosophy*, 13: 505-527.

Buller, D. J. (1999). Defreuding evolutionary psychology: adaptation and human motivation. In Buller, D. J. & Hardcastle, V. G. (eds.), *Where Biology Meets Psychology.* Cambridge, MA: MIT Press.

Buskes, C. (2013). Darwinism extended: a survey of how the idea of cultural evolution evolved. *Philosophia*, 41: 661-691.

Campbell, D. (1974a). Evolutionary epistemology. In Schlipp, P. A. (ed.), *The Philosophy of Karl Popper*. LaSalle: Open Court.

Campbell, D. (1974b). Unjustified variation and selective retention in scientific discovery. In Ayala, F. J. & Dobzhansky. T. (eds.), *Studies in the Philosophy of Biology.* London: Macmillan.

Cavalli-Sforza, L. L. & Feldman, M. W. (1981). *Cultural Transmission and Evolution.* Princeton: Princeton University Press.

Crane, T. (2006). Brentano's concept of intentional inexistence. In Textor, M. (ed.), *The Austrian Contribution to Analytic Philosophy*. New York: Routledge.

Crist, E. (2004). Can an insect speak? The case of the honeybee dance language. *Social Studies of Science, 34* (1), 7-43.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy*, 72: 741-765.

Davies, P. S. (2003). *Norms of Nature: Naturalism and the Nature of Functions*. Cambridge, MA: MIT.

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

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

Dewsbury, D. A. (1999). The proximate and ultimate: past, present, and future. *Behavioral Processes*, 46(3): 189-199.

Dooling, R. and Searcy, M. (1980). Early perceptual selectivity in the swamp sparrow. *Developmental Psychobiology*, 13: 499-506.

Dover, G. (2000). *Dear Mr. Darwin: Letters on the Evolution of Life and Human Nature.* Berkeley, CA: University of California Press.

Dretske, F. (1997). *Naturalizing the Mind.* New York: Bradford.

Dretske, F. (1986). Misrepresentation. In Bogdan, R. (ed.), *Belief: Form, Content and Function*. New York: Oxford University Press.

Ebert, D. (2011). A genome for the environment. *Science*, 331: 539-540.

Ereshefsky, M. (1998). Species pluralism and anti-realism. *Philosophy of Science*, 65: 103-120.

Gallistel, C. R. (1989). Animal cognition: the representation of space, time and number. *Annual Review of Psychology*, 40: 155-189

Gibbard, A. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment.* Cambridge, MA: Harvard University Press.

Godfrey-Smith, P. (1993). Functions: consensus without unity. *Pacific Philosophical Quarterly*, 74: 196-208.

Gould, J. (1975). Honey bee recruitment: the dance-language controversy. *Science, 189* (4204): 685-693.

Grüter, C. et al. (2008). Informational conflicts created by the waggle dance. *Proceedings of the Royal Society*, 275: 1321-1327.

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

Hull, D. L., Langman, R. E., and Glenn, S. S. (2001). A general account of selection: biology, immunology, and behavior. *Behavioral and Brain Sciences*, 24: 511-528.

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

Kornblith, H. (2005). *Knowledge and its Place in Nature.* New York: Oxford University Press.

Kroodsma, D.E. and Miller, E.H. (eds.) (1996) *Ecology and Evolution of Acoustic Communication in Birds*. Ithaca: Cornell University Press.

Laland, K. N. et al. (2011) Is Mayr's proximate/ultimate dichotomy still useful? *Science*, 334: 1512-1516.

Lewis, D. K. (1986). Causal explanation. *Philosophical Papers, Volume 2*. New York: Oxford University Press.

Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics*, 1: 1-18.

Marler, P. and Peters, S. (1988). The role of song phonology and syntax in vocal learning preferences in the songsparrow, *Melospiza melodia*. *Ethology*, 77: 125–149.

Mayr, E. (1988). *Toward a New Philosophy of Biology: Observations of an Evolutionist*. Cambridge, MA: Harvard University Press.

Millikan, R. G. (1984). *Language Thought and Other Biological Categories: New Foundations for Realism.* Cambridge, MA: MIT.

Millikan, R. G. (1986). Thoughts without laws: cognitive science with content. *Philosophical Review*, 95: 47-80.

Millikan, R. G. (1993a). Truth rules, hoverflies, and the Kripke-Wittgenstein paradox. In *White Psychology and Other Essays for Alice*. Cambridge, MA: MIT.

Millikan, R. G. (1993b). Propensities, exaptations, and the brain. In *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT.

Millikan, R. G. (1993c). In defense of proper functions. In *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT.

Millikan, R. G. (1996). On swampkinds. *Mind and Language*, 11(1): 103-117.

Millikan, R. G. (2002). Biofunctions: two paradigms. In Cummins, R. & Ariew, A. & Perlman, M. (eds.), *Function: New Readings in Philosophy of Psychology and Biology*. Oxford: Oxford University Press.

Millikan, R. G. (2004). *Varieties of Meaning: The 2002 Jean Nicod Lectures*. Cambridge, MA: MIT.

Millikan, R. G. (2005). The language-thought partnership: a bird's-eye view. In *Language: A Biological Model*. New York: Oxford University Press.

Millikan, R. G. (2006). Mental content, teleological theories of. *Encyclopedia of Cognitive Science.* DOI: 10.1002/0470018860.s00128

Moczek, A. P. et al. (2011). The role of developmental plasticity in evolutionary innovation. *Proceedings of the Royal Society B, Biological Sciences*, doi:10.1098/rspb.2011.0971.

Mossio, M.; Saborido, C.; Moreno, A. (2009). An organizational account of biological functions. *British Journal for the Philosophy of Science*, 60(4): 813-841.

Neander, K. (1991a). Functions as selected effects: the conceptual analyst's defense. *Philosophy of Science*, 58: 168–184.

Neander, K. (1991b). The teleological notion of 'function'. *Australasian Journal of Philosophy*, 69: 454–468.

Neander, K. (2012). Toward an informational teleosemantics. In Kingsbury, J. & Ryder (eds.), *Millikan and Her Critics.* New York: Wiley-Blackwell.

Nelson, D. A. and Marler, P. (1993). Innate recognition of song in white-crowned sparrows: a role in selective vocal learning? *Animal Behavior*, 46: 806–808.

Okrent, M. (2007). *Rational Animals: The Teleological Roots of Intentionality*. Athens, OH: University of Ohio Press.

Papineau, D. (1986) Representation and explanation. *Philosophy of Science*, 51: 550-572.

Pargetter, R. & Bigelow, R. (1987). Functions. *The Journal of Philosophy*, 84(4): 181-196.

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

Preston, B. (1998). Why is a wing like a spoon? A pluralist theory of functions. *The Journal of Philosophy*, 95: 215-254.

Preston, B. (2009). Biological and cultural proper functions in comparative perspective. In Krohs, U. & Kroes, P. (eds.), *Functions in Biological and Artificial Worlds.* Cambridge, MA: MIT.

Rescorla, M. (2013). Millikan on honeybee navigation and communication. In Ryder et al. (eds.), *Millikan and Her Critics.* New York: Wiley-Blackwell.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Schroeder, T. (2004). New norms of teleosemantics. In Clapin, H. et. al. (eds.), *Representation in Mind: Vol. 1, New Approaches to Mental Representation (Perspectives on Cognitive Science).* New York: Elsevier.

Seager, W. (1999). *Theories of Consciousness: An Introduction and Assessment.* London: Routledge.

Seet, B. et al. (2001). Myopia in Singapore: taking a public health approach. *British Journal* of *Ophthalmology*, 85: 521-526.

Skyrms, Brian (1996). *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Skyrms, B. (2010). *Signals: Evolution, Learning, & Information*. New York, NY: Oxford University Press.

Sober, E. (1984). *The Nature of Selection*. Cambridge, MA: MIT Press.

Stegmann, U. E. (2013). *Animal Communication Theory: Information and Influence*. Cambridge: Cambridge University Press.

Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies* 127: 109-66.

Toulmin, S. (1972). *Human Understanding: The Collective Use and Evolution of Concepts.* Princeton: Princeton University Press.

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

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

Wright, L. (1973). Functions. *The Philosophical Review*, 82: 139–168.

Wright, L. (1976). *Teleological Explanation*. Berkeley, CA: University of California Press.

CHAPTER 2

Evolution, Human Behavior, and Explanation:

Inferential Problems with Evolutionary Psychology

1. Introduction

Evolutionary psychologists believe that they have an inferential strategy that allows them to give accurate evolutionary explanations for contemporary human behavior. In this paper I call the strategy into question, and argue that it is methodologically unsound. The structure of the paper is as follows. In Section Two I discuss the historical and intellectual context from which evolutionary psychology emerged and go on to discuss some of the core theoretical commitments of the discipline. In Section Three, I identify several conditions that must be satisfied in order for evolutionary psychological inferences to go through. In Section Four, I consider the proximate/ultimate distinction and the role of ultimate explanations in evolutionary psychology. In the fifth section, I discuss the problem of individuating behaviors in a way that permits one to make inferences about their ultimate evolutionary functions, and I show that this problem is fatal to the evolutionary psychological enterprise. In Section Six, I consider how the claim that human psychology has been conserved since the EEA ought to be understood. In Section Seven, I illustrate my critique of evolutionary psychology using an example from the literature, and in section eight, I conclude with some final thoughts.

2. The discipline and its core commitments

Explaining human behavior in terms of its biological roots is not new. Darwin treated human behavior, as he did other aspects of organisms, as the result of evolutionary processes, and he forecast that evolutionary theory would one day provide new foundations for psychology.[17] Since Darwin's work, many advocates of evolutionary explanations of human behavior have held (sometimes tacitly) views about human nature which, they believe, are underwritten by evolutionary theory (Young 1985, Segerstralle 2001). Although these views about human nature are not entailed by the theory of evolution, their proponents often presume that they are.

These explanatory endeavors often have a normative dimension: they not only seek to tell us what we are – when they are descriptive – but also, crucially, suggest how we as natural creatures *should* be. They very often endorse the idea that certain innate, biologically-fixed characteristics of our species set limits on what we can become (the idea that certain aspects of human behavior are inevitable or nearly inevitable) and that these characteristics dictate the conditions for human fulfillment – roughly, that we cannot be happy or lead fulfilling lives unless we live in ways that are in some sense natural for us (Antony 2000). One effect of this trend has been to undercut, or at least to displace, explanatory projects in which a greater emphasis is placed on the causal significance of cultural landscapes (Prinz 2012). Those who advocate this approach to explaining human behavior argue that culture sits on a biological platform, and they

---

[17] "In the distant future I see open fields for more important researches. Psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation (Darwin 1859: 424)." See also Richards (2003).

claim that it constrains the degree to which we are psychologically and behaviorally malleable.[18]

The biological approach to explaining human behavior came to fruition during the 1970s, when the discipline of sociobiology was born. Sociobiologists championed evolutionary theory's explanatory power to give *true* origin explanations of the social lives of animals, including human beings. Human sociobiologists focused specifically on human behavior and they believed that their efforts showed that many of our behaviors have their roots in our evolutionary history. They believed that behaviors are brought about by genetic elements and have been subject to natural selection, and that like other genetically based features of organisms, behaviors are inherited. The goal of the project was to demonstrate that behaviors, in this case human behaviors, are largely explicable in terms of the genes which cause them. But worries about the degree to which sociobiology could accommodate facts about behavioral flexibility led to it's being overtaken by the discipline of evolutionary psychology. The charge of this newer approach to human behavior was to focus on the evolutionary functions of the proximate psychological mechanisms which bring behaviors about (Crawford and Krebs 2008). Evolutionary psychologists believe that the cognitive mechanisms responsible for human behavior came into being through evolution by natural selection. More specifically, they claim that the mind is composed of numerous systems called

---

[18] For a powerful corrective to this view, see Henrich et al. (2010).

"cognitive modules" that were designed by natural selection for performing highly specialized tasks: predator avoidance, mate selection, cheater detection, etc.

The claim that the human mind contains specialized modules was first advanced by Fodor (1983), who argued that only input systems – the channels through which sensory input enters the mind – are modular. He argued that mental operations such as belief fixation, inference making, and so on, are not performed by modules, but are performed by domain-general "central" processing.[19] Fodorian modules have nine defining characteristics. They are *informationally encapsulated* (that is, they cannot draw on information residing elsewhere in the mind),[20] they are *inaccessible* to consciousness, they are *mandatory* (that is automatic rather than under conscious control), they process information very *rapidly*, they are *shallow* (that us they use relatively few computational resources), they are *dissociable* (so, damage to a module produces only highly selective, rather than global, cognitive impairment), they are *localizable* in particular neural circuits in the brain, they are *domain specific* (each has a narrow range of inputs), and they are *innate* (in Fodor's [1983] words, they "develop according to specific, endogenously determined patterns under the impact of environmental releasers") (100).

---

[19] It had important precursors in linguistics (Chomsky 1980), systems theory (Simon 1962) and vision science (Marr 1982).

[20] More precisely, "A cognitive system is informationally encapsulated to the extent that the computational operations of the system are insensitive to information from outside the system itself" (Robbins 2012: 642). For Fodor, informational encapsulation is the key element of modularity.

The Fodorian account is a version of "modest modularity." In contrast, evolutionary psychologists claim that the human mind consists *entirely* of modules. This is known as the *massive modularity* hypothesis. Empirical evidence in support of massive modularity is very weak (Buller & Hardcastle 2000, Prinz 2006, Robbins 2013). However, Cosmides and Tooby (1992, 2005) experimentally investigated the psychology of social exchange, and claimed that their findings vindicate the massive modularity hypothesis. They used a modification of what is known as the Wason Selection Task to investigate how we think about deontic rules. The original version of the Wason Selection Task was designed to test how subjects think about material conditionals. They are presented with four cards lying on a table and are told that each has a numeral on one side and a letter on the other. The first card shows an even numeral, the second shows a vowel, the third shows an odd numeral, and the fourth card shows a consonant. Subjects are then asked which two cards need to be turned over to test the claim "If a card has an even numeral showing, then it has a vowel on the other side." Overwhelmingly, the majority of subjects fail to understand that they must turn over the first and fourth cards to test the claim. Based on the assumption that human beings evolved a finely-tuned social intelligence, Cosmides and Tooby predicted that when subjects are asked how to test a *social* rule, one that is a logical equivalent to the material conditional ("if one is under 18, then one is not permitted to drink alcoholic beverages") they would be able to specify what is required. This turned out to be the case. Cosmides and Tooby interpreted this result as showing that the human mind is equipped with a domain-specific "cheater-detection module" that was installed in our

ancestors during the Pleistocene, and which we modern humans have inherited from them. The interpretation of their results has drawn extensive and strong criticisms (e.g., Samuels 1989, Davies et al. 1995, Cheng & Holyoak 1989, Manktelow & Over 1990, Sperber, et al. 1995, Atran 2001, Mallon 2008, Fodor 2008) and it is obvious that their claim about an evolved cheater detection module goes very far beyond the evidence provided by their study.

Given the lack of empirical evidence, it is not surprising that arguments for the massive modularity hypothesis are mainly theory-driven. There are three main theoretical arguments for massive modularity. The one that is most often cited in the literature concerns the adaptiveness of a modular mind. It is sometimes called the "Argument From Design".[21] Recall that evolutionary psychologists believe that human behavior can be understood as being caused by a psychology that was selected for in the past because that psychology was responsive to adaptive challenges. Specialized modules were better at handling such challenges than a general-purpose intelligence would have been. Selection would therefore have favored such an architecture. Therefore, the mind has a modular architecture (Cosmides & Tooby 1992).

What is it that distinguishes evolutionary psychological modules from Fodorian ones? Giving a precise answer to this question is complicated by the fact that there are several notions of modularity found in the literature (Robbins 2013). Fortunately, these fine

---

[21] The other two arguments are sometimes called the "Argument from Evolvability" and the "Argument from Computational Tractability." All three arguments, and objections to them, are nicely set out by Robbins (2009, 2013). See also Carruthers (2006).

distinctions do not make a difference to the arguments in the present paper, so I will limit my remarks to features that all of them share. I have already mentioned that Fodorian modules are peripheral while “Darwinian modules” (as Bermúdez [2005] calls them) are both peripheral and central. This core difference has a crucial theoretical entailment. It forces evolutionary psychologists to abandon the claim that modules are informationally encapsulated (Robbins 2013). They retain all of the other Fodorian properties, but insist that modules are able to draw on information generated by other modules. There are also some restrictions placed on domain specificity. The modules postulated by evolutionary psychologists are supposed to have been sensitive to certain reproductively significant features in the Environment of Evolutionary Adaptedness (EEA). The Environment of Evolutionary Adaptedness consists of the environment(s) in which ancestral organisms faced pressures of living and evolved adaptations in response to them. An animal is said to be “in the EEA” as long as it remains in an environment in which those challenges which gave rise to its phenotypic adaptations persist. In the case of human beings, our present anatomy and physiology were almost entirely in place by approximately 200,000 years ago (Conroy 2005). The environments in which human beings live now are, for almost all of us, importantly different from the environment in which our prehistoric ancestors evolved. According to evolutionary psychologists, although we are no longer in the EEA, our evolved, modular psychology remains adapted to the conditions that obtained in the EEA; “Our modern skulls house a

stone age mind" (Cosmides & Tooby 1997). We have a psychology that is indistinguishable from that of *Homo sapiens* some 200,000 years ago.[22]

It is clear from this that evolutionary psychology is committed to a nativist account of the mind. Since our present-day psychology is constituted by mechanisms which were present a very long time ago, and which have changed very little if at all since that time, contemporary human behaviors are best interpreted along the same lines as their counterparts in the EEA. Evolutionary psychologists conceive of cognitive modules as "very sophisticated computers, whose circuits are elegantly designed to solve the kinds of problems our ancestors routinely faced**....** Behavior in the present is generated by information-processing mechanisms that exist because they solved adaptive problems in the past" (Cosmides & Tooby 1997: 91).

Computational approaches to the mind typically describe mental processes as "software" that is acquired by learning and is run on neural "hardware". But in order for a trait to come under selection, as evolutionary psychologists claim our cognitive information-processing systems did, the cognitive programs must be genetically transmissible. Consequently, *the software has to be built into the hardware* (hard-wired), rather than programmed into it by learning. This is why evolutionary psychologists claim that humans have a deeply rooted psychological organization that is

---

[22] It is worth noting that, although evolutionary psychologists speak about adaptations that arose during the Pleistocene there are features of our nervous system that emerged much earlier. Thanks to Richard Boyd for pointing this out to me. (See also Downes 2010).

embodied in structural features of the brain. A strong sort of nativism is therefore *indispensible* to evolutionary psychologists, and they gladly embrace it. This is made clear in Cosmides and Tooby's (1997) characterization of modules as instinct-like structures. They explicitly endorse William James' account of instincts, and use it to underwrite their conception of how the mind is organized. James (1887) considered instincts to be reflex-like behaviors that "produce certain ends, without foresight of the ends, and without previous education in the performance" (355). Like instincts, modules are said to be innately disposed to bring about certain behaviors. Now, claims that human behaviors are instinctive or instinct-like must accommodate the fact that human behavior exhibits a remarkable degree of plasticity (e.g., Ramachandran 1993, Joblonka & Lamb 2005, Ghalambor, Angeloni, & Carroll 2010, Prinz 2012, Fedyk in press). We are able to meet novel situations in novel ways and to modify our behaviors in response to anticipated outcomes. But such flexibility is hard to reconcile with the claim that human behavior is controlled by instinctual forces. James addressed this challenge by suggesting that behavioral flexibility is explained by our having *very many* instincts, and that a multiplicity of instincts produces the appearance of behavioral flexibility because, in any given circumstance, multiple instincts compete for the control of our behavior. Similarly, Cosmides and Tooby argue that our flexibility is explained by our brains being aggregates of *very many* (hundreds or thousands of) modules that determine human behavior. But Tooby and Cosmides' use of James' account to underwrite a massively modular psychology is not entirely coherent. Notice that James' argument presupposes that instincts are *not* domain-specific. In his view it is only because multiple instincts

compete for the control of behavior that humans have behavioral alternatives. When faced with a danger situation, for example, one might have the instinct to fight or the instinct to flee. Suppose that on any occasion where flight or fight is called for, one or another of these instincts will end up controlling one's behavior. The same cannot be said of *domain-specific* modules. Tooby and Cosmides cannot allow that multiple modules compete for control of behavior in a danger situation because each module is dedicated to processing information about a single domain. So the choice to fight against or flee from an approaching predator has got to be restricted to a *single* module (call it the "predation-avoidance module").[23] If this is the case, then *individual* modules exhibit flexibility, and are therefore unlike Jamesian instincts.

Evolutionary psychologists believe that the theoretical apparatus that I have sketched above provides them with a secure foundation for giving explanations of human behavior, but it is important to distinguish evolutionary psychological explanations of human behavior from evolutionary explanations of human behavior simpliciter. This is particularly important given that evolutionary psychologists often claim that those who reject evolutionary psychology but accept evolutionary theory are committed to a contradiction (they supposedly believe both that the theory of evolution explains the adaptive traits of all biological systems while also denying that evolution explains the

---

[23] This raises an important problem about the individuation of modules. Is there a fight-or-flight module, or does a "fight" module compete with a "flight" module? This is a case of what is known as the "grain problem" (see Sterelny & Griffiths 1992, Atkinson & Wheeler 2003).

configuration of the human mind).[24] It is important to note that evolutionary theory does not *entail* nativism or massive modularity. One might reject the theoretical apparatus proposed by evolutionary psychologists while still embracing an evolutionary account of the human mind. One might, from an evolutionary perspective, regard the mind as a general-purpose learning device that was selected for during the EEA, or as consisting of selected-for modular learning systems that are sensitive to environmental contingencies outside the EEA. One might also endorse the view that the mind is modular without also holding that these modules emerged during the EEA as products of natural selection. It might be that the mind has a modular organization which was acquired ontogenetically (in the lifetimes of individual human beings) rather than phylogenetically (in the lifetime of a taxon). There is, for example, an area of the brain called the "visual word-form area" that is specialized for reading (it is, in effect, a "reading module"). Written language emerged around 3500 years ago (Woods 2010), which is far too recently for reading to have been selected for, suggesting that cognitive modules can be acquired by learning (Dehaene 2009, Dehaene & Cohen 2007, see also Buller & Hardcastle 2000).

Although the massive modularity hypothesis is controversial (Samuels 1998, Fodor 2000, Currie & Sterelny 2000, Buller & Hardcastle 2000, Sterelny 2003, Buller 2005, Prinz 2006), I will not attempt to adjudicate it here. Instead, I will argue that even if it is

---

[24] A prime example is Anne Campbell's frequently quoted remark that such people believe that "evolution stops at the neck" (Campbell 2002:13).

indeed true that our prehistoric ancestors' behavior was underwritten by an evolved modular psychology, and even if it is true that the contemporary human mind has a massively modular organization, *this does not license the sorts of inferences that evolutionary psychologists characteristically make about the psychology of contemporary humans.*

3. The inferential strategy

What is the inferential strategy that evolutionary psychologists use? There are two such. Sometimes they begin by identifying a behavior exhibited by contemporary humans, and set about to identify its evolutionary function. On other occasions they speculate about the sorts of recurrent challenges our prehistoric ancestors faced; they then speculate about how these were responsible for psychological adaptations; and finally, they extrapolate from this to explain features of contemporary psychology. Both kinds of inference rely on practitioners being able to correctly individuate modules. Evolutionary psychologists have to have a way of knowing *which* module causes *which* behavior. Furthermore, since their explanations rest on the claim that the modules that are supposedly hard-wired in human brains are the *very same modules* that were selected for, and that these modules explain both ancestral and contemporary behaviors, it is clear that evolutionary psychologists need to have some method for inferring *which* prehistoric modules are identical to *which* contemporary ones. Unless this can be done it is difficult to see how one can determine the evolutionary functions of contemporary behaviors. However, to the best of my knowledge, there is no discussion about this important methodological point in the literature.

It is a difficult matter to determine whether a trait was under selection, and what it was under selection for doing. In straightforwardly biological cases, scientists use comparative methods, optimality models, and so on, to determine that selection has taken place, and that the items under consideration have retained their ancestral functions (Sober 2008; Orzack & Sober 1994a, 1994b, 1996, 2001). But evolutionary psychologists are faced with special challenges to establishing that the human mind is comprised of many evolved computational structures and that at least some of them have retained their evolutionary functions. For this to happen, at least three things need to be accomplished: 1) one must identify the evolved modules, 2) one must provide independent support for the claim that each module is responsible for the production of certain contemporary behaviors, and 3) one must give evidence for the claim that there are functional, non-trivial similarities between contemporary and ancestral behaviors. The first requirement is necessary because evolutionary psychologists argue that behaviors in the present are caused by cognitive systems which operate today as they did in the past. Recall that each module was supposedly selected for because of its specific fitness enhancing effects. So for example, the mate-procuring module in contemporary humans is so characterized because of its fitness effects in the past. Because it has that specific function it will not be sensitive to situations that require the module that produces behaviors that result in (say) people avoiding poisonous plants. Identifying poisonous plants has fitness-enhancing benefits, but one will not get those benefits unless there is a module that is sensitive to the appropriate range of stimuli. The upshot of this demand is that unless evolutionary psychologists can identify

modules they are not going to be able to say that a particular behavior is underwritten by a specific module with the evolved function of producing behaviors of this sort, and their explanatory project will have difficulty getting off the ground. The second point calls for evidential support. Suppose that we grant evolutionary psychologists the claim that behaviors are caused by special-purpose modules. Such a concession does not necessitate one's accepting their further claims that those modules are 1) the same as ancestral ones, and 2) that contemporary behaviors are caused by them. It is an empirical matter whether the modules that caused behaviors in early humans are the same as those causing behaviors in humans now. Unless this is established, one could claim that special-purpose modules are acquired ontogenetically rather than inherited genetically.

The final point concerns the grounds for matching contemporary modules with ancestral ones. This has got to be based on functional similarity, and this similarity cannot be trivial. The similarity (between a contemporary and an ancestral module) will not be trivial if the function is one that the ancestral module was selected for performing *and* if the contemporary module has the same function *in virtue of its descent from the ancestral module.* This rules out cases where a contemporary module has function *F* due to learning, and this function happens to be the same as that of an ancestral module. It also rules out cases of functional similarity due to convergent evolution. Functional similarity because of selection, which is what evolutionary psychologists want, would likely first require structural similarities between modules, but it is not

obvious that the cognitive architecture that modern humans have is the same as that possessed by our prehistoric ancestors.

All three conditions must be satisfied for the framework to work. It is clear that points 2 and 3 involve *de facto* judgments that contemporary behaviors and psychological phenomena are related to ancestral behaviors and psychological phenomena in a homology-like way in virtue of being underwritten by the same modules. I use the expression "homology-like" because although it is tempting to describe the similarity relations obtaining between ancestral modules and contemporary modules as homologies, they are not homologous in the generally accepted sense of the word. In standard biological usage, "homology"[25] pertains to similarities across taxa in virtue of common ancestry. For example, bird wings and human arms are homologous to the extent that their structural similarities are due to common descent from reptile forelimbs. The sort of similarity that evolutionary psychologists wish to establish is better characterized in terms of similarity due to descent of contemporary phenotypes from ancestral ones. Since homology is standardly understood as a "horizontal" relation (across taxa), I dub the sort of similarity that is the focus of this paper "vertical homology." More specifically, the sort of relation with which I am concerned requires that the function of an ancestral item is conserved over time (there is similarity or

---

[25] I am not changing the meaning of "homology." Rather, I am drawing some finer distinctions to better represent the sorts of relations that are present in evolutionary psychological accounts of human behavior. Although I will use "homology" to keep things simple, it is used to represent similarity within a taxon rather than across taxa.

commonality of function due to descent). Call this *strong vertical homology* to distinguish it from cases where a contemporary item is similar to an ancestral item in virtue of the former's descent from the latter, but without the contemporary item having the same function as the ancestral item.

4. Proximate and ultimate explanations

Giving evolutionary explanations of human behavior engages the distinction between proximate and ultimate explanations, a distinction that was first proposed by Mayr (1961). Proximate explanations pertain to the causal processes that are responsible for the development of organisms,[26] causes operating in fully developed organisms, as well those environmental causes that are external to organisms and which impinge upon them. For example, if I were to offer a proximate explanation of how human kidneys work I might note that kidneys are the body's filtering system and then describe what has to happen for this to occur: that blood gets filtered when the pressure that it exerts causes a cluster of blood vessels to begin the initial process before a tubular structure does the filtering. I might then specify how these structures perform their functions. I might also mention the developmental processes that bring kidneys into being, as well as the role of fluid ingestion, the effects of environmental toxins, and so on. All such explanations gesture at what is going on in a system with kidneys. Without further non-

---

[26] Mayr (1961) suggested that development pertains only to the decoding of "genetic programs." My conception is less restrictive and involves the environment, extra-cellular mechanisms, etc., as has been suggested by Lewontin (2000).

proximate information such explanations provide no guidance about how kidneys came about or what their functions are.

In contrast, ultimate explanations of features of organisms situate those features in an evolutionary context. Such explanations are population-level explanations. So, an ultimate explanation of the human kidney would attribute the initial proliferation of proto-kidneys to the fitness advantages provided by a mechanism[27] for the homeostatic regulation of fluid and solute balance in bony fish, and go on to specify how selection pressures resulted in gradual modifications that eventually gave rise to the mammalian kidney that is able to conserve water while excreting waste. This is an explanation of the *function* of contemporary human kidneys in light of the capacity of ancestral kidneys to enhance the fitness of organisms possessing them.

How does the proximate/ultimate distinction apply to behavior? The first thing to note is that all behaviors are proximately caused. The behaviors of early humans and those of contemporary peoples were and are caused by mechanisms in them and in us. So the issue is whether and how ultimate explanations apply to at least some of our behaviors. To show how the distinction applies in behavioral cases, it is helpful to develop some terminology. I will call behaviors to be explained "target behaviors." One gives a proximate explanation of a target behavior by citing one or more of the causes operating within the organism's lifetime that make a difference to the occurrence of the behavior.

---

[27] This concerns a causal feature of organisms – it's a proximate or set of proximate features which have produced certain effects.

To illustrate this, it is helpful to use an uncontroversial non-human example. Consider biologists' use of the proximate/ultimate distinction to explain the alarm calls made by vervet monkeys. Alarm calls are made in response to the presence of predators, and are proximately caused by factors like perceptions of predators, the causal connection between the perceptions and vocalizations, learning to sound the alarm calls correctly, the developmental processes that underpin their ability to sound the calls, and so forth. A proximate explanation of vervet alarm calls might cite any or all of these factors. Tokens of vervet monkeys' alarm calls are examples of what I shall hereafter call "contemporary target behaviors," and I shall call explanations of them, citing factors like those mentioned above, "proximate explanations of contemporary target behaviors."

Ultimate explanations of target behaviors are explanations of what I will call "ancestral target behaviors" (target behaviors in the EEA). Such explanations will be ultimate explanations of contemporary target behaviors *only on the condition that the ancestral target behavior is identical with (is a strong vertical homolog of) a contemporary target behavior.* Ultimate explanations concern effects of ancestral phenotypes (including behavioral ones) — effects that are taken to have been fitness-enhancing within a certain population in the EEA.[28] It is important to note that these effects are distinct from the proximate causes of the phenotypes which produced them. Consequently,

---

[28] I am assuming, for convenience of expression, that ultimate explanations are adaptationist explanations. While this need not be true, the sort of ultimate explanations of interest to evolutionary psychologists, and hence pertinent to this paper, all involve adaptationist scenarios.

ultimate explanations of features of organisms do not compete with proximate explanations of them.

Consider the alarm call of a contemporary vervet monkey as a target behavior. One offers an ultimate explanation of it by first assuming that the behavior has an evolutionary function. A particular contemporary vervet monkey makes an alarm call because making alarm calls is part of the behavioral repertoire of current members of the vervet population. One further assumes that ancestral vervet monkeys that made and responded appropriately to alarm calls reproduced more successfully than those that did not. This is because the calls correlated enough of the time with the presence of predators and produced appropriate avoidance behaviors in those vervet monkeys that heard them. An additional assumption is that the mechanisms involved in the signaling behavior were passed on genetically to vervet monkey offspring and eventually proliferated through the entire population. This kind of inferential process is involved in giving ultimate explanations, even when the steps are not made explicit.

Science tends to privilege proximate causal explanations (Godfrey-Smith 2003), so one might wonder about the utility of ultimate explanations. If some cognitive mechanism produced a behavior type and the mechanism was selected for because of the effects of the behavior, then one could "black box" its function and focus instead on the interesting proximate causes that produced the behavior. But ultimate explanations do some things that proximate causal explanations do not. They address biological phenomena at the level of *populations* rather than at the level of individuals. They help us make sense of the proliferation of phenotypic traits in populations. Such explanations also

provide an account of what it is for a biological trait to function well, poorly, or not at all. Roughly, an item functions well to the extent that it has the effect that accounted for its proliferation in the EEA (Millikan 1984).

Evolutionary psychologists cannot do without ultimate explanations, but it is doubtful that the explanatory goals towards which they strive are achievable. When biologists give ultimate explanations of nonhuman animal behavior it is generally the case that those animals are still in the EEA. This is true, for example, of the vervet monkey case described above. In such cases, it is trivially true that the behavior under consideration as well as the proximate mechanisms underpinning that behavior are type-identical to a corresponding behavior and the mechanisms that underpinned it in the EEA. But this principle does not apply in the case of human beings. The circumstances of contemporary human life are, in very many respects, quite different from those in which our species evolved. Ultimate explanations of contemporary human target behaviors (and the psychology underpinning them) therefore depend on identifying vertical homologs of those behaviors (and the psychology underpinning them). This is not a straightforward task.

It is important to distinguish ultimate explanations of human behavior and psychology from explanations that only appear to be ultimate ones. I call explanations of the first kind “real ultimate explanations,” and explanations of the second kind “allegedly ultimate explanations.” In what follows, I shall argue that explanations of human behavior of the sort typically given by evolutionary psychologists are allegedly ultimate explanations rather than real ones.

5. Individuating behaviors

Evolutionary psychologists assume that human behaviors, and their underlying modular psychology, came about because of selection for them, and that they have been retained with their original selected-for functions. The upshot is that target behaviors are vertical homologs of ancestral target behaviors in virtue of the fact that the modules that cause them are vertical homologs of ancestral modules that were selected in the EEA for causing just such behaviors. I emphasize these points because as I will show they have deep methodological implications for evolutionary psychology.

Evolutionary psychologists attempt to establish vertical homologies by way of two procedures. One procedure begins by taking a target behavior and identifying its proximate psychological causes: being watchful of a mate might be identified by a psychological state whose content is something like "fearing that" or "believing that" one's mate is cheating. This is what non-evolutionary psychologists do as well: first identify a kind of behavior and then explain that behavior in terms of its proximate psychological causes. Evolutionary psychologists go further. They offer ultimate explanations of those proximate psychological causes by claiming that they have vertical homologs.[29] The other procedure begins with a hypothesis about selection pressures that were encountered by our ancestors in the EEA. In this case, evolutionary psychologists posit hard-wired, ancestral psychological adaptations. They

---

[29] Because psychology can only be fitness-enhancing by producing fitness-enhancing behavioral *effects*, evolutionary psychologists must assume that ancestral psychology produced behaviors of the same sort as contemporary target behaviors.

then go on to propose that this (hypothesized) ancestral psychology is conserved, and therefore (strongly) vertically homologous with the psychology of contemporary human beings. Key point: contemporary peoples' psychology is assumed to be sensitive to the same sorts of inputs, and to produce similar sorts of behavioral outputs, as was the case for our ancestors.

The integrity of this whole explanatory edifice depends on there being criteria for determining which ancestral psychological mechanisms and behaviors are strong vertical homologs of contemporary ones. This issue has been ignored. Instead, the evolutionary psychological enterprise is held together by a tacit assumption that indeed the needed identities have been established. Of course, the fact that this methodological problem has not been addressed does not entail that it cannot be successfully addressed. I will now consider what would be required of evolutionary psychologists in order to secure the inferences they make, and query whether these requirements can possibly be satisfied.

For a contemporary trait to be a strong vertical homolog of an ancestral trait, the contemporary trait must be of the same kind as the ancestral one, it must have the same function as the ancestral one, and must be related by descent to that ancestral trait as part of a reproductive lineage extending back to the EEA. Furthermore, it must be the case that the target trait and the ancestral trait are of the same kind and have the same function *because the target trait is descended from the ancestral trait.* In principle, it might be that a contemporary trait and an ancestral trait are of the same kind and have the same function without one being descended from the other. But if this is

the case, then the contemporary trait is not a vertical homolog of the ancestral one, which would make it impossible to read off an ultimate explanation of the contemporary trait from the ancestral one. This is why the fourth condition, specifying that the sameness relations must depend on descent, is central for evolutionary psychological explanations. It follows from this demanding criterion that evolutionary psychological claims are unfounded unless its practitioners can show that mental modules underpinning present-day behaviors are in fact conserved structures that evolved in the EEA for the performance of adaptive tasks that it is still their function to perform. There are two sorts of considerations that make this especially difficult. The first is epistemic. Psychological mechanisms must be inferred from observations of behaviors. So, knowledge of the mental modules possessed by contemporary humans can only be acquired by making inferences from the behaviors that the modules proximately cause, and knowledge of the modules that populated the minds of our prehistoric ancestors can only be gained by making inferences from the behaviors that they proximately caused. It is worth noting that some evolutionary psychologists would dispute this claim. They argue that we can simply "read off" modules from the adaptive challenges that confronted our prehistoric ancestors (e.g., Buss 1995). For example, if predator-evasion was an adaptive challenge due to the existence of big cats that preyed on humans (as was certainly the case – see Quammen 2004) then natural selection must have seen to it that there was a predator-evasion module. But this strategy cannot work because of what Sterelny and Griffiths (1999) call the "grain problem." Suppose that our prehistoric ancestors had modular minds, and that these modules were indeed

adaptations. Question: was the module that enabled them to avoid being devoured by saber-toothed tigers a saber-toothed tiger avoidance module, a predator-avoidance module, or a danger-avoidance module? As Sterelny and Griffiths perceptively remark, "It is not the existence of a single problem confronting the organism that explains the module, but [assumptions about] the existence of the module that explains why we think of mate choice as a single problem" (1999: 328–329).

Inferring modules from behavior of prehistoric humans is a difficult task since we have only very general evidence of how early humans behaved (see, for example, Kaplan 2002). A more severe problem concerns the causal link between psychological mechanisms and the behaviors that they produce. Evolutionary theory shows that the effects of biological causes may vary as a function of environmental contingencies. Because natural selection concerns the fitness enhancing traits of phenotypes, psychological structures can only be selected if they make a difference to reproductive success by producing behaviors that help spread copies of their genes *in a certain sort of environment*. A structure operating outside the EEA might produce behaviors which promote fitness in ways that are very different from the ways that their prehistoric vertical homologs did, or may even undermine fitness. They might also produce behaviors that differ significantly from the behaviors that they produced in the EEA.[30] Evolutionary psychologists must show that, notwithstanding the confounding effects of

[30] Evolutionary psychologists are often sensitive to this point (e.g., Crawford 1998).

environmental changes, present-day behaviors have their roots in ancestral homologs that had a positive effect on the fitness of humans in the EEA.

Even if it were possible to establish that contemporary target behaviors are produced by evolved cognitive modules, there would still be obstacles to identifying these with *particular kinds* of behaviors and modules hypothesized to have existed in the EEA. But this is precisely what is required for ultimate explanations of the sort given by evolutionary psychologists to succeed. Modules can be individuated only by the behaviors that they produce, so their individuation is parasitic on the individuation of behaviors. We ordinarily individuate behaviors by citing agents' intentions (a behavior counts as answering the telephone if it was performed with the intention of answering the telephone). Evolutionary psychologists cannot use this method for individuating behaviors, because they offer *subpersonal* explanations for the production of behavior, and do not address how, if at all, subpersonal explanations can be brought into relation to personal level explanations.[31] Saying that a behavior is of a certain type in virtue of the intentions that produced it does not allow one to infer a subpersonal module responsible for the behavior (modules are subpersonal computational mechanisms, not rational agents with beliefs and desires).

There are three options available for individuating behaviors. One is to individuate them by their effects, another is to individuate them by their functions, and yet a third is to

---

[31] See Bermúdez (2005) for a thorough discussion of this problem.

individuate them by their causes. I will address each of these and show that evolutionary psychologists cannot appeal to any of them.

The first option is to individuate behaviors by their effects. Evolutionary psychologists might claim that a contemporary target behavior is the same kind as an ancestral behavior only if both target and ancestral behaviors have the same kind of effects. Now, consider the claim that a target behavior is the same kind as a behavior in the EEA. The first point to notice is that when one is dealing with claims about contemporary and ancestral behaviors there is always an epistemic asymmetry at work, since there is no way of observing what early humans did and what effects resulted from their doings. Of course, paleoanthropologists can and do make inferences about prehistoric human behavior based on material culture and forensic evidence, but these inferences are far too coarse-grained to be of service to evolutionary psychologists. Consequently, they have to resort to *speculations* about what happened, which weakens the authority of their claims. In most cases, it is not possible to determine the effects of prehistoric behaviors, but even if the epistemic obstacles can be surmounted, serious difficulties remain. Suppose that two behaviors are tokens of the same type if they have effects of the same type. The problem remains of establishing common causes for the two behaviors. Suppose that an ancestral behavior and a contemporary behavior have the same kind of effect. This would not license the inference that the psychological processes responsible for the target behavior were also responsible for the ancestral behavior. So, individuating behaviors by their effects does not allow one to infer that contemporary target behaviors are underpinned by conserved modules.

The second option is to individuate behaviors by their functions. By this criterion a target behavior is identical with an ancestral behavior if the two behaviors share the same function. This approach has a strong intuitive appeal. Hunting with stones and hunting with guns both count as hunting because both have the function of bringing down game. This approach is not vulnerable to the difficulty of inferring modules from behaviors, but it suffers from a much deeper problem of circularity. Individuating a target behavior by its function is the same as offering an ultimate explanation of that behavior. But if one begins with the assumption that a target behavior has a function *one has already presupposed that this very behavior was selected for in the EEA*. In other words, one supposes that a behavior was selected for and then uses this supposition as evidence that the behavior was selected for. A further problem with this strategy is that evolutionary psychologists hold that it is psychological mechanisms rather than behaviors that are selected-for. Behaviors have selected functions only derivatively, as expressions of modules that were under selection in the EEA. So the claim that some behavior has a certain function rests on the assumption that the behavior is produced by a module that was selected for performing this function. This brings us back to the original problem of making inferences from behaviors to modules.

Although the function of a phenotypic trait is seen in the effects that it produces, individuating behaviors by their functions does not collapse into individuating them by their effects. The function of a phenotypic trait is the effect of that trait on *fitness* in a *critical mass* of cases in the EEA. In any particular case, tokens of the trait might fail to produce the fitness-enhancing effect. Consider two birds of the same species, both of

which perform courtship displays. One bird's display has the effect of attracting a mate, but the other's does not. If what makes the two behaviors fall under the category "courtship display" is their function, it is clear that effects of a behavior need not always accord with its function. In this case two different effects are produced and it is not clear if only one or both is a selected function

Finally, one might individuate behaviors by their causes. This seems to be the only option evolutionary psychologists, because they need to be able to infer underlying psychology from behavioral effects in order to give ultimate explanations of present-day psychology. On this option, if two behaviors have the same (proximate) psychological causes then they belong to the same behavioral kind. Now, consider the claim that a target behavior is the same kind as a behavior in the EEA. If behaviors are individuated by their causes, then the contemporary target behavior that one wishes to explain, and the behavior in the EEA by means of which one wishes to explain it, must have the same kind of causes. This strategy too is circular because it is proposed that cognitive modules can only be individuated by the behaviors that they bring about, and it is also proposed that these behaviors are individuated by the modules that cause them.

I have already pointed out that if an organism is still in the EEA, or is in an environment that is relevantly similar to the EEA, one can offer real ultimate explanations of its behavior and psychology. The reason has to do with the relationship between genetic developmental programs and the environments in which the programs operate. Very often, environments are implicitly regarded as causally inefficacious backgrounds to development (West-Eberhard 2003), but this is misleading, because even subtle

environmental variations can lead to striking phenotypic differences. For example, water fleas (*Daphnia cucullata*) that develop in an environment where there are chemicals indicating the presence of predators, undergo dramatic morphological changes. They develop impressive, helmet-like structures on their necks and spines along their tails. These are epigenetic effects that persist across generations (Tolliran and Dodson 1999, Agrawell, Laforsch, and Tolliran 1999), and there are no such effects in environments in which these chemicals are absent. Environmentally-induced developmental changes affecting behavior have been observed in a variety of species. For example, infant Bonnet macaques that are raised in conditions where the efforts required for forging food are highly variable (sometimes requiring extensive foraging and sometimes not) are strikingly more timid than those raised in conditions where efforts required for forging are more consistent (conditions in which the macaques either do not ever have to forage extensively or always have to forage extensively). The infants raised in environments with variable foraging demands also show signs of depression of the sort normally observed only in maternally deprived primates, and as adolescents they are less inclined to engage in social play behavior. Recent studies demonstrate that the inconsistent form of provisioning alters the development of the monkeys' neural systems that mediate responses to stress (Copland *et al.* 1996, 1998).

One reason for the widespread neglect of the causal significance of environments for development may have to do with the tacit assumption that environments have not changed in important ways since the EEA. This assumption licenses unwarranted claims about selection. When we say that some feature of an organism was selected

for, this is really a shorthand expression for a complex causal story involving genetics, developmental pathways, and chemical signaling systems, the result of all of which is differential fitness relative to an environment. *All* of these processes are sensitive to environmental contingencies. The environment-relative nature of the processes responsible for producing phenotypes is such that even if *the very same causal mechanisms* underpin the behavior of an ancestral individual and a present-day individual, these might produce divergent behavioral effects. For this reason (as well as others), one cannot simply infer sameness of psychological causes from sameness of behavioral effects. Even if a contemporary target behavior is the output of a mental module, it does not follow that the same module would have produced the same behavior it produced in the EEA. The causal role of environments weakens the explanatory glue between evolved psychological mechanisms and their behavioral effects, and casts doubt on the claim that if a certain behavior is reliably produced by a certain causal mechanism in a present-day environment that is far removed from the circumstances in which our prehistoric ancestors lived, then the presence of that same behavior (however individuated) permits one to infer that the behavior stemmed from the same cause in the EEA.

One might deal with this problem by "black boxing" the causes of ancestral behaviors and saying something like "selection resulted in developmental processes that gave rise to such–and-such behaviors" while remaining agnostic about the nature of these developmental processes. Clearly, early humans did have adaptive behavioral phenotypes and there were developmental processes at work that brought these into

being. However, this black-boxing strategy is not available to evolutionary psychologists, because they regard their project as one of *discovering* the deep structure of the mind and thereby giving ultimate explanations of present-day psychology. Tooby and Cosmides put this very clearly, writing that, thanks to evolutionary psychological research, "in 50 or 100 years one will be able to pick up an equivalent reference work [to Gray's Anatomy] for psychology and find in it detailed information-processing descriptions of the multitude of evolved species-typical adaptations of the human mind'' (1992: 69).

For organisms that are still in the EEA, one can infer that the developmental factors have remained more or less the same since the time when they were initially selected. In such cases, the proximate causes of a contemporary target behavior can reasonably be thought to coincide with the proximate causes of that behavior earlier in the lineage. Because the environment has remained constant in these cases, it is possible to "read back" the underlying causal processes to say something about their nature in ancestral populations.[32] However, ultimate explanations of human behaviors are not normally of this sort, as many aspects of our environments have changed in significant ways since the EEA. There are special limitations that apply in such cases. One cannot justifiably "read back" the proximate causes of the behavior into the EEA because changes in the

---

[32] Even in these cases the inference will be unreliable if, at some point in the past, the population has encountered a bottleneck. If a population undergoes a drastic collapse, it may be that the genetic profile of the pre-bottleneck population differs significantly from the post-bottleneck population, and consequently that the causal underpinnings of phenotypic traits differ as well (Richard Boyd, personal communication).

environment may have altered the behavioral effects of proximate psychological causes. So it looks like knowing that a certain contemporary target behavior is caused by some proximate psychological process *does not allow one to assume that this process was selected for in the EEA as the cause of some corresponding ancestral behavior*. One cannot establish that a contemporary target behavior is caused in the same way as a similar ancestral behavior because we have no access to the psychological causes of the ancestral behavior (as we have seen, evolutionary psychologists cannot rely on individuating behaviors by their causes).

What emerges from all of this is that ultimate explanations of human behaviors in the EEA may shed no light on contemporary target behaviors. If ancestral psychology *has* been conserved since the Pleistocene, it might have manifestations in contemporary behaviors that are quite different from those that the same psychology produced in ancestral environments. The key worry in trying to individuate behaviors by their causes in the context of ultimate explanations concerns the order of explanation. To say of any two behaviors that they are the same because they have causes of the same kind, one first has to establish that the psychological processes that cause the behaviors are the same. It is only *after* this has been established that one can conclude that the two behaviors are the same. But this forecloses the possibility of inferring the psychological cause of a prehistoric behavior from the psychological cause of a current behavior, as evolutionary psychologists try to do.

The preceding discussion has shown that even if it were possible to know that a contemporary target behavior is produced by a module that is strongly vertically

homologous to some module that was selected for during the EEA, there is no way to infer *which* ancestral module is homologous with the contemporary one. This shows that ultimate explanations of the sort that are offered by evolutionary psychologists are not scientifically viable.

6. What is conserved?

The considerations that I have presented prompt thoughts on the uses of the notion of "conservation" as applied to behavior. Genes, anatomical structures, physiological processes, and behaviors are said to be conserved if they persist in a lineage and "highly conserved" if they persist in a lineage despite speciation. Suppose that it is true that that there is some sense in which human psychology has been conserved since the Pleistocene or before. This might be understood as a claim that the *causal mechanisms* that gave rise to a behavior in the EEA have been conserved without the behavior that they produced in the EEA having been conserved. Alternatively, it might be meant as a claim that behaviors prevalent in the EEA have been conserved without their causal mechanisms having been conserved. Or it might be meant as a claim that both causal mechanisms and behaviors have been conserved. To say that a *behavior* has been conserved is only to say that a target behavior is "the same" as some ancestral behavior by some criterion or other. This weak sense of "conservation" cannot support evolutionary psychological claims because it says nothing about the underlying psychology. So, in order for their project to go through, evolutionary psychologists must assert that the causal mechanisms underpinning the target behavior have been conserved. To justify this, they need an individuation criterion that allows them to infer

sameness of psychological cause from sameness of behavioral effects, *both now and in the EEA*. In other words, the pattern of reasoning that would allow one to offer ultimate explanations for the psychological causes of target behaviors depends on both psychological causes and behavioral effects being conserved.

7. An example of evolutionary psychology in action

I turn now to an example of what is known as the bottom-up strategy (Buss 2004) to illustrate how evolutionary psychology falls foul of the individuation problem. I will show how this leads practitioners to offer allegedly ultimate explanations of contemporary target behaviors in place of genuine ones. It is important to emphasize that this is not a cherry-picked example. The methodological deficits that it so clearly illustrates are rampant in the evolutionary psychological literature. I have selected the example because it displays these deficits especially clearly.

The study, entitled "Sex differences in perceptions of infidelity: men often assume the worst" (Goetz & Causey 2009) appeared in *Evolutionary Psychology*, a well-regarded peer-reviewed electronic journal.

The authors begin with a claim about conditions obtaining in the EEA. Drawing on Trivers' (1972) Parental Investment Theory, they state that it would have been more biologically costly for men than for women to fail to detect a partner's infidelity.

> Ancestral men…were susceptible to an additional and profound cost if they failed to detect a partner's infidelity: cuckoldry—the unwitting investment of resources into genetically unrelated offspring. Cuckoldry was one of the most serious

> threats to fitness our male ancestors faced. Some of the costs associated with cuckoldry include misdirection of the male's time, effort, and recourses to rearing a rival's offspring, loss of time, effort, and resources the man spent attracting his partner, and reputational damage if such information becomes known to others (255)

The claims made here about the social and sexual behavior of prehistoric humans far exceed anything warranted by paleoanthropological evidence. But we can ignore this shortcoming because even if these claims were well founded, the inference made from them would still be unwarranted. Suppose that these claims about early humans are true. The authors infer that they "provided selection pressure for an arsenal of anti-cuckoldry tactics in men" and that one of these may have been "evolved psychological mechanisms designed to overperceive the likelihood of their partner's infidelity." On the assumption that a mate's sexual infidelity was more costly for ancestral men than it was for ancestral women, the authors infer that contemporary men should be more suspicious of their partner's future infidelity than contemporary women are.

> Due to the costs associated with being cuckolded, men's infidelity detection system may have been designed to overestimate the likelihood of their partner's future infidelity. This overestimation bias would have generated behavior aimed at preventing infidelity, such as increased vigilance, mate guarding, and even affectionate behavior (258).

The study proceeded as follows. One hundred sixty-three male and female college students were asked (1) "How likely do you think it is that you will in the future have

sexual intercourse with someone other than your current partner?" and (2) "Please indicate your agreement or disagreement with the following statement: 'I will probably be sexually unfaithful to my partner.'" They were also asked (3) "How likely do you think it is that your current partner will in the future have sexual intercourse with someone other than you, while in a relationship with you?" and (4) "Please indicate your agreement or disagreement with the following statement: 'My partner will probably be sexually unfaithful to me in the future.'"

> Using two independent samples, two different response formats, and two data collection methods, we found support for our hypothesis. Men's perceptions of the likelihood of their partner's future infidelities were greater than women's. In conclusion, we found support for the hypothesis that men's infidelity detection system should be designed to overestimate the likelihood of their partner's infidelity (262).

Now, consider the inferential structure of this study. The authors begin with a theory-driven conjecture that in the EEA mate infidelity was more costly for males than it was for females, and that those males who were good at safeguarding against their mates' infidelity would have been more fit than those who were not able to guard against it as effectively. They infer from this that selection favored a tendency for males to be more suspicious of their mates than females are, and that this explains *hypothesized* mate guarding behavior in the EEA. These *ancestral target behaviors* (increased vigilance, mate guarding, and even affectionate behavior) are then *ultimately explained* as having the function of preventing their mates from clandestinely conceiving offspring with other

men in virtue of having been underwritten by selected-for cognitive mechanisms (responsible for the tendency to "overperceive" infidelity). Finally, the responses to the investigators' questions about partners' future infidelity are the *contemporary target behavior.* This behavior is assumed to be a strong vertical homolog of the ancestral target behavior, and therefore being amenable to the same ultimate explanation.

Confronted with evidence that 21st century male American college students[33] are more doubtful of the future sexual fidelity of their mates than their female counterparts are, the authors *assume* that (a) male college students' expressed skepticism is caused by a hardwired, domain-specific cognitive module, and (b) that the *same* module existed in Pleistocene males, and that it produced behaviors of the same sort as the contemporary behavior. These unsound assumptions are supposed to underwrite the conclusion that the sexual suspiciousness of contemporary males is caused by an innate cognitive mechanism with the evolutionary function of enhancing their fitness by preventing cuckoldry. But the authors do not provide support for their claim that the psychological mechanism driving contemporary male sexual skepticism is the very same mechanism that (supposedly) drove prehistoric anti-cuckoldry behavior. In light of the arguments that I have presented, it is not possible to provide such support. Consequently, the authors' inferences about the evolutionary roots of the male students attitudes – their

---

[33] The tendency by psychologists to ignore the remarkable degree to which culture influences psychology, and to draw sweeping conclusions about human beings on the basis of samples drawn from Western, educated, industrialized, rich, and democratic societies is powerfully argued by Henrich et al. (2010).

allegedly ultimate explanation of it – is unjustified. It might be true, but we have not been given good reasons to accept it as true. Additionally, the authors help themselves to assumptions about the computational structure of the ancestral module – namely, that it produced skepticism about the fidelity of mates. This claim is not entailed by the hypothesis that there was selection in the EEA for a module with the function of guarding against female infidelity. Even if there were good evidence that the sexual skepticism of contemporary males is underwritten by a domain-specific module, this would fall short of showing that it is plausible that the contemporary module is a strong vertical homolog of an ancestral module with the function of preventing infidelity.

8. Conclusion

Evolutionary psychologists' claims about ancestral modules are formulated on theoretical grounds. They suppose that recurring adaptive challenges were likely to give rise to mental adaptations in the EEA, and then suppose that these modules underpin contemporary behavior. Suppose that one could establish, on these sorts of theoretical grounds, that the minds of our Pleistocene ancestors possessed a module that was responsible for a certain sort of domain-specific behavior. For example, suppose that one could establish that ancestral females possessed a "mate-selection module" — a mental system that was sensitive to whatever attributes of potential mates were correlated (in the EEA) with reproductive value, and which regulated mating behaviors performed by these females. Now, what conclusions about the mate-selection module would this claim license? Nothing other than that there was a module dedicated to regulating female mating behavior in the EEA. On its own, it would not

license conclusions about the inputs to which the module was tuned, nor would it license conclusions about the behaviors that it brought about. Given that the evolutionary hypothesis does not provide any specific information about the computational structure of the ancestral module, we cannot extrapolate from it to draw conclusions about the psychological mechanisms that regulate the mating behavior of contemporary women. Even if one knew that the mating behavior of contemporary women is regulated by a conserved module, the evolutionary hypothesis would not underwrite the inference that this module is the same one that regulated the mating behavior of ancestral females. All of this prompts the question, "is evolutionary psychology possible?"

David Buller (2005) thinks that it is possible to give contentful explanations of human behavior situated in the context of evolutionary theory. He distinguishes evolutionary psychology from what he calls Evolutionary Psychology.

> The former is a field of inquiry, a loose confederation of research programs that vary widely in theoretical and methodological commitments and that are federated only by a commitment to ''adopting an evolutionary perspective on human behavior and psychology'' (Barrett et al. 2002: 1). The latter, Evolutionary Psychology, is a specific doctrinaire research program within this field of inquiry, a central doctrine of which is the so-called massive modularity hypothesis (MMH) (881).

If the arguments presented in this paper succeed, the methodological defects of Evolutionary Psychology (as opposed to evolutionary psychology) are so severe as to

be unrectifiable, and consequently Evolutionary Psychology is not viable. Evolutionary Psychologists simply do not have the methodological resources to justify the claim that the psychological causes of contemporary behaviors are strong vertical homologs of the psychological causes of corresponding behaviors in the EEA. The verdict for evolutionary psychology (as opposed to Evolutionary Psychology) is less clear. It should go without saying that the human mind is a product of evolution, and evolution must therefore enter into an explanation of human psychology in some way. Evolutionary Psychology rests on three pillars: the massive modularity hypothesis, the claim that modules evolved as adaptations to recurrent challenges in the EEA, and the tacit assumption that modules can be individuated and so license claims about strong vertical homologies. These three components, taken together, are inconsistent with the competing evolutionary hypothesis that evolution fashioned the human mind as a domain-general or modestly modular learning system. On this account, the architecture of the human mind (whatever that turns out to be) was selected to be adaptive and malleable, rather than fixed and instinct-like, and supports a view of "human nature" that is far less reductive and nativist than the version that is promulgated by Evolutionary Psychologists (as well as sociobiologists and other social scientists influenced by evolutionary thinking). Importantly, this competing hypothesis is immune from the criticisms that I have developed in this paper, chiefly because it does not reduce the mind to an array of domain-specific systems and require that these are homologs of ancestral systems. It is also well-supported by recent work in developmental biology indicating that it is often the case that the behavioral repertoire of even invertebrates is

often highly malleable and driven by learning (West-Eberhart 2003, Menzel & Benjamin 2013). However, a research program of this sort – one which restricts itself to scientifically justifiable claims about the phylogenetic roots of human psychology, and which gives developmental plasticity and learning their due – is unlikely to have much utility for explaining the *specifics* of human behavior biologically. It is likely to be less contentful than Evolutionary Psychology precisely because it makes no attempt to extend biological explanations to domains where they are not of service.

REFERENCES

Agrawal A. A., Laforsch C., & Tollrian R. (1999). Transgenerational induction of defences in animals and plants. *Nature,* 401: 60-63.

Antony L. M. (2000). Norms and natures. *Ethics*, 111(1): 8-36.

Atkinson A. P. & Wheeler M. (2003). Evolutionary psychology's grain problem and the cognitive neuroscience of reasoning. In Over, D. E. (ed.), *Evolution and the Psychology of Thinking: The Debate.* Hove, Sussex: Psychology Press.

Atran, S. (2001). A cheater-detection module? Dubious interpretations of the Wason selection task and logic. *Evolution and Cognition*, 7(2): 1-7.

Barrett, L. et al. (2002). *Human Evolutionary Psychology*. Princeton: Princeton University Press.

Bermúdez, J. L. (2005). *Philosophy of Psychology: A Contemporary Introduction*. New York: Routledge.

Buller, D. J. (2005). *Adapting Minds: Evolutionary Psychology and the Persistant Quest for Human Nature.* Cambridge: MIT.

Buller, D. J. & Hardcastle, V. G. (2000). Evolutionary psychology, meet developmental neurobiology: against promiscuous modularity. *Brain and Mind*, 1: 307-325.

Buss D. M. (1995). Evolutionary psychology: a new paradigm for psychological science. *Psychological Inquiry*, 6:1-30.

Buss, D. M. (2004). *Evolutionary Psychology: The New Science of the Mind*. Boston: Pearson.

Campbell, A. (2002). *A Mind of Her Own: The Evolutionary Psychology of Women*. Oxford: Oxford University Press.

Cheng, P.W. & Holyoak, K.J. (1989). On the natural selection of reasoning theories. *Cognition,* 33: 285-313.

Chomsky N. (1980) Rules and representations. *Behavioral and Brain Sciences,* 3:1–15.

Crawford, C. & Krebs, D (2008). *Foundations of Evolutionary Psychology.* New York: Lawrence Erlbaum Associates.

Coplan J. D., *et. al.* (1996). Persistent elevations of cerebrospinal fluid concentrations of corticotropin-releasing factor in adult nonhuman primates exposed to early-life stressors: implications for the pathophysiology of mood and anxiety disorders. *Proceedings of the National Academy of Sciences, USA*, 93(4): 1619-1623.

Coplan J. D., et al. (1998). Cerebrospinal fluid concentrations of somatostatin and biogenic amines in grown primates reared by mothers exposed to manipulated foraging conditions. *Archives of General Psychiatry*, 55(5): 473-477.

Conroy, G. C. (2005). *Reconstructing Human Origins*. New York: W. W. Norton and Company.

Cosmides, L. & Tooby, J. (1997). Evolutionary psychology: a primer. In Downes, S. & Machery, E. (eds.), *Arguing About Human Nature: Contemporary Debates*. New York: Routledge, 2013.

Cosmides, L. & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange." In Buss, D. (ed.), *The Handbook of Evolutionary Psychology.* Hoboken, NJ: Wiley.

Crawford, C. (1998). The theory of evolution in the study of human behavior: an introduction and overview. In Crawford, C. & Krebs, D. (eds.), *A Handbook of Evolutionary psychology: Ideas, Issues, Applications.* Mahwah, N.J.: Lawrence Erlbaum Associates.

Currie, G. & Sterelny, K. (2000). How to think about the modularity of mind-reading. *Philosophical Quarterly*, 50: 145-160.

Davies, P. S.; Fetzer, J. H. & Foster, T. R. (1995). Logical reasoning and domain specificity. *Biology and Philosophy***,** 10 (1): 1-37.

Dehaene, S. (2009). *Reading in the Brain.* New York: Viking.

Dehaene, S. & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15 (6): 254-262.

Downes, S. M. (2010). The basic components of the human mind were not solidified during the Pleistocene epoch. In Ayala, F. J. & Arp, R. (eds.), *Contemporary Debates in Philosophy of Biology*. New York: Wiley-Blackwell.

Fedyk, M. (forthcoming). How (not) to bring psychology and biology together. *Philosophical Studies.*

Fodor, J. A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press.

Fodor, J. A. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology.* Cambridge, MA: MIT Press.

Fodor, J. A. (2008). Comment on Cosmides and Tooby. In Sinnott-Armstrong, W. (ed.), *Moral Psychology: The Evolution of Morality: Adaptations and Innateness* (*Moral Psychology*, volume 1). Cambridge, MA: MIT Press.

Ghalambor, C. K, Angeloni, L, & Carroll, S. P. (2010). Behavior as phenotypic plasticity. In Westneat, D. & Fox, C. W. (eds.), *Behavioral Ecology*. Oxford University Press.

Godfrey-Smith, P. (2003). *Theory and Reality: An Introduction to the Philosophy of Science.* Chicago: University of Chicago Press.

Goetz, D. & Causey, K. (2009). Sex differences in perceptions of infidelity: men often assume the worst. *Evolutionary Psychology*, 7(2): 253-263.

Henrich, J. et al. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33: 61-135.

James, W. (1887). What is an instinct? *Scribner's Magazine*, 1(3): 355-366.

Kaplan J.M. (2002). Historical evidence and human adaptations. *Philosophy of Science*, 69: 294-304.

Laland, K. N. & Brown, G. R. (2011). *Sense and Nonsense: Evolutionary Perspectives on Human Behavior.* New York: Oxford University Press.

Lewontin, R. C. (2000). *The Triple Helix: Gene, Organism, and Environment.* Cambridge, MA: Harvard University Press.

Mallon, R. (2008). Ought we to abandon a domain general treatment of 'ought'?" In Sinnott-Armstrong, W. (ed.), *Moral Psychology, The Evolution of Morality: Adaptations and Innateness*, (*Moral Psychology*, *Vol. 1*), Cambridge, MA: MIT Press.

Manktelow K.I. & Over D.E. (1990). Deontic thought and the selection task. In: Gilhooly K.J.et al. (eds.), *Lines of Thinking: Reflections on the Psychology of Thought: Representation, Reasoning, Analogy and Decision Making*. New York: Wiley.

Marr D. (1982). *Vision*. San Francisco: W. H. Freeman.

Mayr, E. (1961). Cause and effect in biology. *Science*, 131: 1501-1506.

Mayr, E. (1983). How to carry out the adaptationist program? *The American Naturalist*, 121: 324-334.

Menzel, R. & Benjamin, P. R. (2013). *Invertebrate Learning and Memory (Handbook of Behavioral Neuroscience, Vol. 22*). New York: Elsevier.

Orzack, S. H. & Sober, E. (1994a). Optimality models and the test of adaptationism. *The American Naturalist*, 143: 361-380.

Orzack, S. H. & Sober, E. (1994b). How (not) to test an optimality model. *Trends in Ecology and Evolution*, 9: 265–267.

Orzack, S. H. & Sober, E. (1996). How to formulate and test adaptationism. *The American Naturalist*, 148: 202-210.

Orzack, S. H. & Sober, E. (2001). *Adaptationism and Optimality.* New York: Cambridge University Press.

Pinker, S. (2002). *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking.

Prinz, J. J. (2006). Is the mind really modular? In Stainton, R. (ed.), *Contemporary Debates in Cognitive Science*. Oxford: Blackwell.

Prinz, J. J. (2012). *Beyond Human Nature: How Culture and Experience Shape the Human Mind*. New York: W. W. Norton & Co.

Quammen, D. (2004). *Monster of God: The Man-Eating Predator in the Jungles of History and the Mind*. New York: W. W. Norton & Co.

Ramachandran, V. S. (1993). Behavioral and magnetoencephalographic correlates of plasticity in the adult human brain. *Proceedings of the National Academy of Sciences, USA*, 90: 10413-10420.

Richards, R. J. (2003). Darwin on mind, morals, and emotion. In Hodge, J. & Radick, G. (eds.), *The Cambridge Companion to Darwin.* Cambridge: Cambridge University Press.

Robbins, P. (2009). Modularity of Mind. *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2010/entries/modularity-mind/>.

Robbins, P. (2013). Modularity and mental architecture. *Wiley Reviews in Cognitive Science*, 4: 641-649.

Samuels, R. (1998). Evolutionary psychology and the massive modularity hypothesis. *British Journal for the Philosophy of Science,* 49: 575-602.

Samuels, R. (2000). Massively modular minds: evolutionary psychology and cognitive architecture. In Carruthers, P. & Chamberlain, A. (eds.), *Evolution and the Human Mind*. New York: Cambridge University Press.

Segerstralle, U. (2001). *Defenders of the Truth: The Battle for Science in the Sociobiology Debate and Beyond*. New York: Oxford University Press.

Simon H. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society,* 106:467-482.

Sperber D. et al. (1995). Relevance theory explains the selection task. *Cognition* 57: 31-95.

Sterelny, K. (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. Oxford: Blackwell.

Sterelny K. & Griffiths P.E. (1999). *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: University of Chicago Press.

Tolliran R. & Dodson S. I. (1999). Inducible defenses in Cladocera: constraints, costs and multipredator environments. In Tolliran R & Harvell C. D. (eds.), *The Ecology and Evolution of Inducible Defenses*. Princteton, NJ: Princeton University Press.

Tooby, J. & Cosmides, L. (1992). In Barkow *et al.* (eds.), *Cognitive Adaptations for Social Exchange*. New York: Oxford University Press.

Tooby, J. & Cosmides, L. (1995). Foreword. In Baron-Cohen, S., *Mindblindness: An Essay on Autism and Theory of Mind.* Cambridge, MA: MIT Press.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology,* 46: 35-57.

Trivers, R. L. (1972). Parental investment and sexual selection. In Campbell, B. (ed.), *Sexual Selection and the Descent of Man*, 1871-1971. Chicago, IL: Aldine.

Young, R. M. (1985). *Darwin's Metaphor: Nature's Place in Victorian Culture.* Cambridge: Cambridge University Press.

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

Woods, C. (2010). *Visible Language: Inventions of Writing in the Ancient Middle East and Beyond*. Chicago: The Oriental Institute.

CHAPTER 3

So Many Ways to be Wrong about Evolution:

The Strange Case of Joyce's Evolutionary Debunking Argument

1. Introduction

A number of accounts have been put forth in support of the view that morality evolved (e.g., Gibbard 1992, Katz 2000, Sinnott-Armstrong 2007, Fleming & Levinson 2012). Proponents of such accounts utilize evolutionary theory to make substantive claims about the origin and, most importantly, the functions of morality. For some, the evolution of morality has implications for metaethics (e.g., Richards 1986, Campbell 1996, Casebeer 2003, Street 2006). In this Chapter I am going to discuss Richard Joyce's foray into the evolution of morality, as set out in his book of the same name (2006). Joyce offers a rich account of what he believes the evolution of morality to have consisted in, and argues that this evolutionary story undermines moral realism. The standard use of evolutionary theory, both in biology and in the social sciences, is to account for the origin and function of phenotypic features. Joyce uses it to give an account of the origin and function of moral judgment, but he also uses evolutionary theory to craft a skeptical argument to the effect that moral beliefs are never epistemically justified, and therefore that moral realism is probably false.

Joyce's argument is an example of what have come to be called "evolutionary debunking arguments" (Kahane 2011). These are arguments that purport to show that

the evolutionary genealogy of certain sorts of belief (typically, evaluative beliefs) undermines the likelihood of their being true, but which are structured in a way that avoids the genetic fallacy. There are two parts to such arguments. The first part is empirical. It consists of descriptive premises which lead to the conclusion that a certain sort of belief has an evolutionary etiology. The second part is epistemic. It involves premises which lead to the conclusion that the etiological account gives one reason not to be confident in the truth of beliefs of that kind. The general form of such arguments, then, can be captured in the following formula: *if such-and-such an evolutionary explanation of a class of beliefs is true, then this shows that the beliefs under consideration are unjustified.* The evolutionary facts about the etiology of the beliefs are said to *exclude* the possibility that they are justified. The idea is this: there is an inverse relation between the strength (plausibility) of the evolutionary story and the degree to which one should have confident in the beliefs that it purports to explain.

On the face of it, then, such arguments will be philosophically significant only to the degree that the empirical claims on which their skeptical conclusions are founded are likely to be true. I say "on the face of it" because I will later show that the resources of evolutionary theory are insufficient to warrant Joyce's conclusion, and that he must in the end invoke conceptual issues.

Joyce's account of how morality evolved leans heavily on evolutionary psychology, both in form and in substance. Like his fellow debunker Sharon Street (2006) he holds that evolutionary psychology has substantive implications for philosophy. He believes that evolutionary psychological explanations can rule out certain philosophical positions and

underwrite others. As I have articulated in Chapter Two of this dissertation, evolutionary psychology is methodologically flawed. Consequently, Joyce's use of it is riddled with difficulties. In the course of my discussion, I will concentrate on three of difficulties that his argument encounters. These are summarized as follows. I argued in Chapter Two that evolutionary psychological inferences do not address the problem of establishing that present-day psychological mechanisms are function-preserving reproductions of ancestral psychological mechanisms. In the present chapter I will show that this worry applies in equal measure to Joyce's use of evolutionary psychology. I will argue that he has not established – or come anywhere near to establishing – that the proximate moral psychology of contemporary human beings is the *very same* moral psychology that was selected for in the EEA. His failure to secure this identity undermines the credibility of the empirical portion of his argument and therefore, by extension, the credibility of the metaethical conclusion of the argument as a whole. My second concern also pertains to the empirical portion of his argument. I will show that although Joyce claims that his evolutionary speculations are supported empirically, his justificatory strategy is flawed. I argue that he demonstrates only that his evolutionary story is *consistent* with certain data. Consistency is important, because a claim that is not consistent with the relevant data is likely to be false. But mere consistency is not enough. For one's claim to be empirically supported it must be more plausible than rival claims that are also consistent with the data. Joyce fails to deliver on this requirement. My third criticism concerns the inference from the empirical portion of the argument to its metaethical conclusion. I will demonstrate that Joyce's metaethical conclusion does not follow from his empirical

premises. Even if all of Joyce's empirical considerations are sound, the metaethical lesson that he draws from them is unjustified unless he introduces an additional, highly implausible premise. Along the way I will touch upon other weaknesses in a more perfunctory manner.

The organization of the chapter is as follows. In Section Two, I will set out Joyce's account of the evolution of the moral sense. To do this, I will explain his view of what it is that distinguished proto-moral attitudes from true moral judgment, and then set out his thesis that both were selected for in the EEA because of their reproductive benefits. In Section Three, I will explore the grounds for Joyce's choice of just this evolutionary hypothesis. I will show that, like all such accounts, it rests on questionable assumptions about the conservation of psychological phenotypes. I will then argue that the empirical considerations that he cites do not provide significant support for his evolutionary hypothesis. In Section Four I will explain Joyce's argument in support of the metaethical entailments of his evolutionary story, and then, in Section Five, I will show why, even if the empirical portion of the argument is sound, his metaethical conclusions are unwarranted.

2. Joyce's evolutionary hypothesis

The claim that morality evolved is ambiguous. It might mean that we *Homo sapiens* are disposed to be kind, generous, cooperative, intolerant of injustice, and so on, because evolution shaped us to behave in these ways. Notice that this does not entail anything about the nature of the psychological mechanisms that are responsible for these

dispositions. Alternatively, the claim that morality evolved might be a claim about moral psychology. For Joyce, being a moral animal is not just a matter of how one behaves; it is a matter of why one behaves that way. Moral behavior is behavior that is, at least in part, motivated by moral judgments. Joyce holds that the capacity for moral judgment proceeds from a psychological faculty which he calls the moral sense.

Inclinations to be helpful and aversions to being harmful are not in themselves moral. They are the pre-moral "building blocks" (de Waal 2009) of morality. It is generally accepted that prosocial behaviors were selected for in social species (e.g., Wilson 1975, Axelrod & Hamilton 1981, Maynard Smith 1982), because cooperation enhances genetic fitness. However, on Joyce's account, instinctive prosociality was not sufficient for securing cooperation among early humans. Human beings evolved cognitive capacities that endowed them with tremendous behavioral flexibility. Thanks to our capacity for instrumental reasoning, we are able to choose courses of action that benefit us as individuals to the detriment of the fitness of our genes. Consequently, the innate prosocial dispositions of early human beings could be trumped by self-interested motives. Our ancestors' flexible psychology made them fickle. Given the fitness benefits of cooperation and the power of self-interest to undermine it, there was (Joyce speculates) selection pressure for the emergence of a psychological faculty for safeguarding prosocial behavior. This is how the moral sense came into being.

The moral sense has several distinctive features. One of them is domain specificity. The moral sense is attuned to the social sphere. It is "a specialized mechanism functioning to govern an adaptive behavior" (114). Thus, the moral sense is very like, if

not identical to, what evolutionary psychologists call a "cognitive module" (see Chapter Two). A second important feature of the moral sense is that it operates with linguistically-infused concepts such as "desert" and "transgression," which are required for distinctively moral attitudes such as guilt and blame. This implies that language had to have become established before morality could evolve, and given that morality involves sophisticated concepts, hominin language abilities would have had to have been correspondingly sophisticated for this to happen.[34] Third, it is central to Joyce's conception that moral judgments have practical clout. They have practical clout because of their perceived authority and inescapability. The authority of a moral judgment is its binding force. Moral judgments are felt to take precedence over prudential or hedonic considerations that conflict with them. The inescapability of moral judgments is driven by the fact that they are represented not as subjective evaluations of the world, but as objective properties of it. From a subjective perspective, it is not just that one morally disapproves of x. Rather, it is that one disapproves of x because x has the property of moral badness. It is the practical clout of moral judgments that put pressure on our prehistoric ancestors to conform to and enforce norms of cooperation. And the claim is that it exerts the same kind of pressure on the behavior of modern human beings. Consequently, "Moral judgment can thus function as a kind of social

---

[34] There is a great deal of scientific controversy about when it was that language emerged (Christiansen & Kirby 2004). If language emerged fairly recently, this may falsify Joyce's hypothesis about the moral sense, because a special-purpose, biologically unprecedented cognitive module would have taken quite a long time to evolve.

glue, bonding individuals together in a shared justificatory structure and providing a tool for solving many group coordination problems" (Joyce 2006: 117).

3. The proximate mechanism

Joyce's empirical hypothesis is that moral judgments are outputs of a domain-specific cognitive mechanism that was selected for stabilizing prosocial behavior by buffering it against the effects of self-interested motives. Now, it is a truism that proximate mechanisms are underdetermined by selection pressures. That is, for any set of selection pressures, there are a number of proximate mechanisms that might serve as candidates for responding to those pressures. Perhaps an example will make this clear. Animals that rely extensively on vision are vulnerable to predation at night, because it is difficult for them to detect, and therefore to evade, animals that prey on them. Among nocturnal animals, those individuals that are able to detect the presence of predators in conditions of poor illumination (all things being equal) have greater reproductive success than those that are unable to do this. For nocturnal animals living in ecologies where nocturnal predators are abundant, there is a selection pressure for traits that protect them from nocturnal predation. That is to say, under these circumstances, animals best able to avoid nocturnal predation will leave more descendants than their less well-endowed conspecifics. How might selection operate in such circumstances? There are a number of possibilities. There might be selection for more acute night vision, or for enhanced auditory abilities, or for the ability to sense the heat of a nearby predator's body, or for an odor that repels predators, or for greater locomotor speed, or for silent locomotion, or for any number of other phenotypic variations. This example

makes clear that knowing that certain selection pressures obtained in the EEA does not license inferences about which proximate mechanisms were selected for responding to those pressures.

Suppose that Joyce's hypothesis that there was a selection pressure for stabilizing prosocial dispositions is true. Evolution might have settled upon any of a number of "solutions" to this problem. If there was some adaptive response to these selection pressures, all that one needs to assume is that there was selection for *some* proximate mechanism that stabilized prosocial attitudes (Boyd unpublished manuscript). Why should one suppose that Joyce has identified the proximate mechanism that was *in fact* selected for? That is, why should one think that Joyce's hypothesis is more plausible than its projectable alternatives? Why not suppose that there was selection for stronger prosocial emotions or greater indoctrinability instead of selection for a moral sense? Joyce is not entirely unaware of the problem, but he misconstrues it in a way that underestimates its force for his project. This is clearly shown by his response to David Lahti's criticism. Lahti (2003) pointed out that it seems odd that natural selection would create a biologically novel mechanism for regulating cooperative behavior instead of strengthening prehistoric humans' desire to cooperate. Joyce's reply turns on an analogy with other proximate mechanisms.

> Think…about the psychological reward systems that have evolved in humans regarding sex and eating. One might ask why natural selection bothered giving us all that complicated physiological equipment needed for having an orgasm – why not design us simply to want to have sex? Natural selection did make us

> want to have sex, and one of the means of securing this desire was precisely the human orgasm….And perhaps natural selection has made us want to cooperate, and granting us a tendency to think of cooperation in moral terms (where this includes the capacity for guilt) is a means of securing this desire (114-115).

This response misses the point of Lahti's criticism entirely. The point at issue is not whether the moral sense is a desire-strengthening faculty. Lahti's point concerns the grounds for supposing that evolutionary processes gave rise to an entirely new faculty rather than modifying an existing one. Lahti's worry is grounded in a heuristic that biologists use to decide among alternative hypotheses. He is invoking the principle that evolution "tinkers" with existing structures. Over time, the cumulative effects of many such changes add up to major phenotypic modifications. Lahti's point is that taking into account the gradualistic character of natural selection allows one to limit the "possibility space" of candidate proximate mechanisms.

Now, returning to the question of Joyce's motivation for settling on this particular proximate mechanism, one might suggest that it was guided by his antirealist commitments. I think a more likely explanation is that he has lost sight of the methodological importance of entertaining projectable alternatives when offering empirical explanations (as is often the case in evolutionary psychological accounts). Recall that Joyce is trying to give an explanation of a feature of *contemporary moral psychology* – namely, the capacity for moral judgment. He is not just extrapolating from an evolutionary scenario. Instead, he has a contemporary explanatory target in mind from which he infers an evolutionary etiology. Suppose that one grants that Joyce is

correct in saying that contemporary people experience moral judgment as authoritative, inescapable, and as involving the existence of objective moral facts (this seems more than just plausible). Further, suppose that the function of this faculty (the moral sense) *in contemporary humans* is for stabilizing prosocial dispositions (again, a very plausible assumption). Now, what do these facts imply? On Joyce's view these facts are best explained by the assumption that the moral sense evolved and that it was conserved along with its ancestral function. But why think that an evolutionary story is the best explanation of the moral sense in contemporary people? To justify his evolutionary turn, Joyce must show that it is implausible that the moral sense was acquired by means of social learning, and he attempts to do this in the following way. He holds that if the moral sense was selected for then it must be innate. This entails that if the moral sense is *not* innate, then it was not selected for. If one accepts this, then the reasonable thing to do is to evaluate the relevant scientific literature to determine whether the evidence favors a nativist account of moral judgment. Joyce derives five major points from this literature in favor of moral nativism. These include the universality of morality among humans, the broad similarity of moral norms across far-flung cultures, children's development of morality in the absence of explicit instruction, children's ability to distinguish moral norms from conventional norms, and children's facility handling deontic rules. Joyce concedes that none of these considerations decisively support nativism, but he argues that they count in favor of it.[35] Here, Joyce is using the same

---

[35] Yet another methodological problem is that Joyce does not undertake a comparable

top-down explanatory strategy that is often adopted by evolutionary psychologists: first identify a contemporary psychological or behavioral trait, next argue that it is innate, and finally propose an evolutionary scenario that would explain why the trait is (innately) present in contemporary humans.

4. Two problems with Joyce's empirical argument

Nowhere are the methodological shortcomings of Joyce's approach more obvious than in his attempt to neutralize the charge that his evolutionary account is nothing more than a just-so story. Gould and Lewontin (1979) introduced the term "just-so story" to ridicule speculative adaptationist accounts of human behavior that were at the time rampant in the sociobiological literature. They were however not very clear about exactly what qualifies an evolutionary hypothesis as a just-so story. As I understand it, just-so stories are adaptationist hypotheses about contemporary traits that meet only the lowest standard of scientific justification: they are *possibly true explanations*. An explanation is possibly true just in case it is consistent with whatever it is supposed to explain. Now, for anything that one wishes to explain there is an indefinitely large pool of possibly explanations, not all of which will be plausible. Among the plausible ones, some will be more plausible than others (on grounds such as parsimony, consistency with well-confirmed hypotheses, successful predictions entailed by it, etc.). The plausibility of an

---

survey of empirical evidence supporting an anti-nativist account. This makes it impossible for the reader to form an evaluation of the strength of the evidence in favor of nativism. For anti-nativist interpretations of the evidence presented by Joyce, see Prinz (2008).

explanation, relative to other possible explanations, is roughly its degree of empirical support. So, a just-so story is an evolutionary explanation that is consistent with the phenomenon to be explained but which is no more plausible than other possible candidates. Now, consider Joyce's defense of his position in light of these distinctions. He writes, "I am not putting this hypothesis forward as *true*. It is a hypothesis that is plausible, coherent, and testable – and its truth remains to be established. However, there is good reason for looking favorably upon it" (134). If there is good (empirical) reason for looking favorably upon a hypothesis it is, by my lights, a plausible hypothesis. So, Joyce seems to be using "plausible" in much the same way as I intend it. And if his hypothesis is plausible (in this sense) then it is not a just-so story. So, to evaluate Joyce's defense, it is necessary to consider how and to what degree it is supported by empirical considerations. I have already noted that the hypothesis is consistent with a range of empirical evidence, so it is possibly true. What is needed now is something stronger, something that indicates that the hypothesis is preferable, on empirical grounds, to other possibly true explanations. But Joyce does not supply any such evidence. Instead, he seems to think that evidence in favor of moral nativism is sufficient to establish the plausibility of his theory. So the theory remains vulnerable to the accusation that it is a just-so story. There is nothing in his account that should lead one to prefer it to rival explanations of moral nativism (recall that insufficient evidence has been given to warrant moral nativism over moral non-nativism – see footnote 2).

There is also a deeper problem handicapping Joyce's argument. The problem is this. Joyce wishes to bring two things together under the umbrella of a unified explanation.

One is a set of selection pressures in the EEA, and the other is a present-day psychological capacity. Joyce wants to claim that the present-day psychological capacity is the very same proximate mechanism that was selected for in the EEA, but he has no means of establishing their identity. Because human beings are no longer in the EEA, one cannot simply state of some psychological trait that it was selected for doing what it now does. Because psychological traits do not leave traces in the fossil record, one cannot use paleontological evidence to establish structural continuity between ancestral minds and modern minds. And because the moral sense is, by definition, unique to humans one cannot use comparative methods to establish that it is highly conserved (unlike the neurological reward systems for sex and eating that Joyce offered in response to Lahti's criticism). Putting the point somewhat differently, Joyce needs to motivate the claim that the moral sense is identical to whatever proximate mechanism emerged in response to the pressures that he supposes to have been operative in the EEA. Using the terminology developed in Chapter Two, Joyce does not establish that the moral sense is a strong vertical homolog of some proximate mechanism that was selected for stabilizing prosocial attitudes in the EEA. He has therefore offered an allegedly ultimate explanation of the moral sense rather than a real ultimate explanation of it.

5. The inference to moral skepticism

At the beginning of Joyce's (2008) précis of *The Evolution of Morality* he characterizes his project as attempting to accomplish two tasks.

> The first is to clarify and provisionally advocate the thesis that human morality is a distinct adaptation wrought by biological natural selection. The second is to inquire whether this empirical thesis would, if true, have any metaethical implications (213).

In the preceding sections of this chapter I have demonstrated that the arguments that he uses to accomplish the first task are multiply flawed. I have shown that he has not supplied good reasons for accepting, even provisionally, his evolutionary hypothesis. Given the structure of the argument, this is sufficient to undermine its metaethical conclusion. However, argumentative flaws are not always fatal. Suppose that the criticisms that I have leveled against the empirical components of the argument can be met. What then? In the remainder of this chapter I wish to argue that there are further difficulties with the argument. These difficulties go beyond worries about its empirical components. I will argue that even if one ignores the weaknesses of the empirical phase of the argument, and treats the empirical premises as sound, the skeptical metaethical conclusion does not follow from them.

Joyce's inference to moral skepticism proceeds as follows. Suppose that the moral sense can be given an exhaustively evolutionary adaptationist explanation. To explain some biological feature as an adaptation is to make the case that the feature is such that it made the organisms that possessed it more fit, as measured in reproductive success, than those that did not possess it. So, the claim that the moral sense was an adaptation is the claim that those ancestral humans that had the moral sense were more fit and had greater reproductive success than those that lacked it. Because,

according to Joyce, the moral sense is genetically fixed, the descendants of early humans also possess it. Hence, the moral sense came to be all but universal in our species. In order to explain how this occurred, one needs to explain how the fitness-enhancing effects of the moral sense were mediated. That is, one needs to explain what it was about having the capacity for forming moral beliefs that resulted in those organisms have greater reproductive success than members of the population that lacked that capacity. There was something about that capacity that gave them a reproductive edge. There are two broad possibilities. One is that, for any selected-for belief-forming capacity, one might explain the effect on fitness by citing the truth of the beliefs so formed. Suppose that there is a cognitive faculty that was selected for forming beliefs about biological species.[36] It seems reasonable that any fitness benefits that might have been accrued from such a faculty can be accounted for by its producing true beliefs about the plants and animals in the EEA. This is because prehistoric hominins would have been unable to survive if they could not form such true beliefs. However, if Joyce's story is accurate, then the fitness benefits that were bestowed on early humans in virtue of their possession of a moral sense had nothing to do with the truth of moral beliefs. These benefits can be exhaustively explained by the stabilizing effect of the moral sense on prosocial behavior. So, there is no reason to think that any moral beliefs are true. As Joyce puts the point:

---

[36] Some ethnobiologists hold that this is the case (Medin & Atran 1999).

> We have an empirically confirmed theory about where our moral judgments come from (we are supposing). This theory [i] doesn't state or imply that they are true, [ii] it doesn't have a background assumption that they are true, and, importantly, [iii] their truth is not surreptitiously buried in the theory by virtue of any form of moral naturalism. This amounts to the discovery that our moral beliefs are the product of a process that is entirely independent of their truth, which forces the recognition that we have no grounds one way or the other for maintaining those beliefs. (211)

Of course, certain moral beliefs intuitively strike one as being true. But on Joyce's view this is because evolution has deceptively fashioned the human mind to misrepresent morally evaluative assessments as objective facts. Once one understands the circumstances in which the moral sense emerged, and the function that it served in those circumstances, it becomes possible to expose the belief in moral facts as an illusion that was foisted upon us by evolution.

Before going on to explain what is wrong with this argument, it is necessary to pause to consider Joyce's point (iii). He introduces this proviso because if moral properties can be brought into line with natural properties (by virtue of being identical to them, constituting them, being realized by them, or some other relation) then it might be possible that there are facts that answer to moral judgments. Suppose that the case can be made that moral goodness is realized by cooperativeness, kindness, reciprocity, and so on. These are all characteristics that are, according to the debunking argument, stabilized by the moral sense. Consequently, the moral sense stabilizes behaviors and

attitudes that are *good*. In order for his argument to terminate in moral skepticism Joyce must close the door to this possibility. He does this by offering a version of Mackie's (1977) Argument from Queerness. Moral demands, he claims, are binding and non-negotiable. They seem to demand conformity to the exclusion of anything else (for example, hedonic or prudential considerations). Moral naturalists must, Joyce believes, show how natural properties can have practical clout. Roughly, whatever natural properties goodness consists in must be such that they are intrinsically prescriptive. Anyone acquainted with a moral fact would thereby be obligated to act in accordance with that fact. There are no such facts, so moral naturalism is false.

There are several responses that the realist can make to this sort of argument. Two of these situate practical clout in the minds of the people making judgments, rather than in the things being judged. One response is to say that it is the fact that one *believes* an item to be good that supplies the judgment with practical clout. Another response is to argue that moral beliefs are causally inert and that moral desires supply practical clout. On these views, moral judgments are inescapable and authoritative for much the same reason that other beliefs or desires have practical clout (Brink 1984). A third option is to reject Joyce's presumption that moral realism must come wrapped up in an objectivist package. Objectivists hold that moral facts must be mind-independent – one's attitude must play no part in making it the case that something is morally good or bad. Subjectivism rejects this presumption and has it that moral claims are true or false in virtue of facts about the mind of the person doing the judging. Joyce does not address the subjectivist challenge (see Kahane 2011).

In attempting to undermine moral naturalism Joyce is faced with a dilemma. He can either show that moral naturalism is false, or he cannot. If he can show that it is false (as he thinks he has), then the debunking argument is not needed, and the evolutionary speculations are superfluous. And if he cannot show that moral naturalism is false (as its defended believe), then the debunking argument cannot go through. So, either the debunking argument is unnecessary or it cannot go through. What is revealed here is that Joyce's evolutionary speculations *play no real role in his inference to moral skepticism*. Non-empirical, conceptual considerations are doing all of the philosophical work.

6. An antidote to the Napoleon pill

Having raised a major skeptical doubt about the professed role of claims about the evolutionary etiology of moral judgments in Joyce's theory, I will set it aside. My final task in this chapter is to show that Joyce's skeptical conclusion is not justified even if one accepts all of his prior claims.

Joyce's fictional example of a Napoleon pill is a useful point of entry for my argument.[37] Suppose that there is a pill which, if swallowed, causes one to have beliefs involving the concept *Napoleon* (call these "Napoleon beliefs"). The pill also causes one to have amnesia about having taken it. Further, suppose that people who do not take the pill

---

[37] There are actually two Napoleon pill examples in *The Evolution of Morality*. I am using the second one, which conforms more precisely to the structure of the debunking argument.

never have Napoleon beliefs because the pill is necessary for having Napoleon beliefs. You discover that you have taken the pill. Joyce argues, plausibly enough, that learning that you have taken the pill ought to undermine all of your Napoleon beliefs, because you come to understand that your Napoleon beliefs were caused by a truth-insensitive process. You should, he thinks, take an antidote to rid yourself of the epistemically suspect beliefs. Having done this, if you are interested in the warrant for Napoleon beliefs, you should seek out reliable sources of evidence about the existence or non-existence of Napoleon.[38] Joyce argues, analogously, that if there is a truth-insensitive evolutionary process that causes one to have moral beliefs this removes any warrant that one might have for believing that there are moral facts. "If the analogy is reasonable," he adds, "…it would appear that once we become aware of this genealogy of morals we should (epistemically) do something analogous to taking the antidote pill: cultivate agnosticism concerning all positive beliefs involving these concepts until we find some positive evidence for or against them" (2006: 181).

To pinpoint what is wrong with this argument, consider that taking the pill is necessary for forming Napoleon beliefs. The evolutionary analog of this would be to claim that were it not for the evolution of the moral sense nobody would have moral beliefs (beliefs featuring moral concepts). But this is clearly wrong, as is evidenced by the fact that Joyce has written a whole book setting out his views about moral concepts using his

---

[38] Notice that this flies in the face of the previous claim that the pill is *necessary* for forming Napoleon beliefs because it is necessary for having the concept *Napoleon*.

rational, conscious, domain-general intelligence rather than a domain-specific module for making moral judgments. So, the Napoleon pill story is importantly disanalogous to the evolutionary case. The evolutionary story does not require that moral concepts *must* emanate from the moral sense, or even that the moral sense is necessary for making judgments that involve moral concepts. The evolutionary hypothesis requires only that the moral sense evolved for making moral judgments. Importantly, it does not show that moral judgments cannot also be made by other cognitive systems, including the domain-general one. This implies that if moral facts exist, they might be known by means of rational reflection or some other reliable domain-general cognitive process. Suppose that this is the case. Presumably, if domain-general intelligence was selected for, it was because of the fitness benefits that rationality provided to our ancestors. These benefits would be best accounted for by the truth-tracking power of rationality. It follows that if rational deliberation can yield moral judgments these judgments are not vulnerable to the evolutionary debunking strategy. There are two responses available to the evolutionary debunker, both of which are unsatisfactory. The first is to insist that although moral judgments can be made on the basis of rational deliberation, all such judgments are erroneous because there are no moral facts. The problem with this response is that it begs the question of moral properties in favor of the moral skeptic. If moral realism is demonstrably unwarranted, and this can be shown on non-evolutionary grounds, then it is not clear why one would introduce an evolutionary argument for establishing that establishing that moral realism is unwarranted. The other option is to refashion the evolutionary story to bring it in line with the Napoleon pill analogy. One

might do this by stipulating that the moral sense is the only source of moral judgments. This option is problematic in several ways. First, it flies in the face of empirical research indicating that people reach moral conclusions by the exercise of conscious reflection as well as through automatic, unconscious processing (e.g., Cushman et al. 2006, Pizarro & Bloom 2003, Paxton & Greene 2010).[39] Second, it contradicts introspective awareness of rational moral deliberation. And third, it is not warranted by the evolutionary theory that Joyce endorses. The fact that some feature of an organism was selected for performing a certain function does not exclude the possibility that that function can be performed by a different feature of the organism that was not selected for performing it.

7. Conclusion

Richard Joyce's debunking argument is an ambitious attempt to use evolutionary psychology to undermine moral realism. If I am right, and I think that I am, he has failed to achieve his ambition. Joyce failed, in part, because of his use of evolutionary psychology to frame his argument. In doing this, he imported methodological errors that are endemic to that discipline (e.g. Dupré 2001, Buller 2005). He does not seem to appreciate that one cannot simply "read off" proximate mechanisms from selection pressures. That ancestral organisms were faced with certain adaptive challenges does not entail anything about *which* (if any) proximate mechanisms were selected for

---

[39] These are known as "dual process" models of moral cognition.

responding to those challenges. This oversight leads Joyce to move far too quickly from his speculative evolutionary scenario to the claim that the moral sense emerged in response to these pressures. His failure to consider what other proximate mechanisms might have been at work makes his account a just-so story. Looking more closely, it seems that Joyce's single-minded focus on the moral sense can be explained by his use of a top-down inferential strategy rather than an extrapolative one. He begins with a psychological phenomenon found in contemporary humans and argues that certain selection pressures assumed to have been operative in the EEA likely accounted for its emergence. Joyce's adaptationist stance may explain why his discussion of the empirical literature is so perfunctory, and why he neglects to consider evidence that favors anti-nativism. Evolutionary stories that are constructed in this way do not show that contemporary psychological mechanisms are the *very same mechanisms* as those that emerged in the EEA. In common with evolutionary psychologists, Joyce seems to be unaware of the problem.

Joyce's third misappropriation of evolutionary theory concerns his unarticulated (and obviously unjustified) assumption that if a proximate mechanism is selected for performing function, then no other proximate mechanism can perform function. Without this assumption, he cannot argue that the etiology of the moral sense excludes the possibility that moral beliefs track moral truths, and the debunking strategy cannot go through.

That the failure of Joyce's argument can be laid at the doorstep of evolutionary psychology suggests that any philosophical enterprise that uncritically draws on this

discipline may be hampered by its methodological shortcomings. But if the problem lies in how evolutionary psychology is *presently* done, might it not be possible to purge it of its errors and to use the purified version for philosophical purposes? In my view, the prospects for this are dim, because evolutionary psychology *of any sort* invites one to do the impossible. By its very nature it requires one to make claims about selection pressures that cannot be substantiated, to make unfalsifiable claims about the proximate mechanisms that were selected for responding to those pressures, and to make unjustified assumptions about the relations that obtain between features of our contemporary psychology and features of the psychology of prehistoric hominins. Evolutionary psychology is speculative through and through, and it is therefore important that philosophers who wish to incorporate evolutionary biology into their research program resist its seductive appeal.

REFERENCES

Axelrod, R. & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211: 1390-1396.

Boyd, R. N. (unpublished manuscript). On being 'Oh so scientific': evolutionary theory as methodological anesthesia in moral philosophy and moral psychology.

Brink, D. O. (1984). Moral realism and the sceptical arguments from disagreement and queerness. *Australasian Journal of Philosophy*, 62(2): 111-125.

Buller, D. (2005). *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature.* Cambridge, MA: MIT Press.

Campbell, R. (1996). Can biology make ethics objective? *Biology and Philosophy*, 11: 21-31.

Casebeer, W. D. (2003). *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition.* Cambridge, MA: MIT Press.

Cushman, F. et al. (2006). The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychological Science,* 17(12): 1082-1089.

Dupre, J. (2001). *Human Nature and the Limits of Science.* Oxford: Clarendon Press.

Fleming, J. E. & Levinson, S. (2012). *Evolution and Morality (Nomos LII).* New York: NYU Press.

Gibbard, A. (1992). *Wise Choices, Apt Feelings: A Theory of Normative Judgment.* Cambridge, MA: Harvard University Press.

Gould, S. J. & Lewontion, R. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme*. Proceedings of the Royal Society of London, Series B,* 205(1161): 581-598.

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

Joyce, R. (2008). Précis of *The Evolution of Morality*. *Philosophy and Phenomenological Research*, 77(1): 213-218.

Kahane, G. (2011). Evolutionary debunking arguments. *Noûs, 45(1): 103-125.*

Katz, L. D. (2000). *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives.* New York: Imprint Academic.

Lahti, D. C. (2003). Parting with illusions in evolutionary ethics. *Biology and Philosophy,* 18: 639-651.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong.* Harmondsworth: Penguin.

Maynard Smith, J. (1982). *Evolution and the Theory of Games.* Cambridge: Cambridge University Press.

Medin, D. & Atran, S. (1999). *Folkbiology.* Cambridge, MA: MIT Press.

Paxton, J. M. & Greene, J. D. (2010). Moral reasoning: hints and allegations. *Topics in Cognitive Science,* 2(3): 1-17.

Pizarro, D. A. & Bloom, P. (2003). The intelligence of the moral intuitions: comment on Haidt (2001). *Psychological Review*, 110: 193-198.

Prinz, J. (2008). Acquired moral truths. *Philosophy and Phenomenological Research*, 77(1): 219-227.

Richards, R. J. (1986). A defense of evolutionary ethics. *Biology and Philosophy,* 1: 265-293.

Sinnott-Armstrong, W. (2007). *Moral Psychology, The Evolution of Morality: Adaptations and Innateness, Vol. 1.* Cambridge, MA: MIT Press.

Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies*, 127: 109-166.

de Waal, F. (2009). *Primates and Philosophers: How Morality Evolved*. Princeton, NJ: Princeton University Press.

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