Darwinian Humility:
Epistemological Applications of Evolutionary Science

by
Said Saillant
B.A. Philosophy, Psychology, Cognitive Science
Rutgers University-New Brunswick, 2011

Submitted to the Department of Linguistics and Philosophy
in Partial Fulfilment of the Requirements for the Degree of
Doctor of Philosophy in Philosophy
at the
Massachusetts Institute of Technology
September 2017

Signature redacted

Signature of Author. .....
Department of Linguistics and Philosophy
September 8, 2017

Signature redacted

Certified by..... .....
Roger White
Professor of Philosophy
Thesis Supervisor

Signature redacted

Accepted by..... .....
Roger White
Professor of Philosophy
Chair of the Committee on Graduate Students

Darwinian Humility:
Epistemological Applications of Evolutionary Science

by
Said Saillant

Submitted to the Department of Linguistics and Philosophy on September 8, 2017
in Partial Fulfillment of the Requirements for the Degree of Doctor of
Philosophy in Philosophy.

ABSTRACT

I use evolutionary science – its tenets and theory, as well as the evidence for it – to investigate the extent and nature of human knowledge by exploring the relation between human cognition, epistemic luck, and biological and cultural fitness. In "The Epistemic Upshot of Adaptationist Explanation," I argue that knowledge of the evolution by natural selection of human cognition might either defeat, bolster, or preclude the epistemic justification of our current beliefs. In "The Evolutionary Challenge and the Evolutionary Debunking of Morality," I argue that we lack the evidence to know whether human moral knowledge evolved or exists. In "Human Morality: Lie or Heirloom?," I argue that, contrary to the popular conception of their descent, human moral belief systems might ultimately be the result of ancient parental deception.

The project unfolds against the backdrop of Darwinian naturalism, that all living beings on Earth are related by descent with modification and that natural selection has been the main (but not exclusive) means of modification. The central lesson is that human knowledge attribution is more epistemically demanding than previously thought because to self-ascribe knowledge with justification we must justify the assumption that certain unconfirmed evolutionary hypotheses are correct. The ultimate hope is to give epistemology a Darwinian update and, in consequence, human knowledge its proper place in nature.

Thesis Supervisor: Roger White
Title: Professor of Philosophy

Contents

Acknowledgements

Many people have influenced the ideas developed in this dissertation. First, I owe thanks to my wonderful dissertation committee: Roger White, Kieran Setiya, Philip Kitcher, and Alex Byrne. Second, I would like to thank my fellow graduate students, especially Dylan Bianchi, Nilanjan Das, Sophie Horowitz, Brendan de Kennessey, Bernhard Salow, and Ian Wells, for valuable conversations that have shaped my interests over the years. Third, for helpful comments, discussion, questions, and suggestions, thanks to Dan Baras, Selim Berker, Sylvain Bromberger, Justin Clarke-Doane, Tom Donaldson, Ryan Doody, Kevin Dorst, Nicole Dular, David Enoch, Cosmo Grant, Lyndal Grant, Caspar Hare, Sally Haslanger, Jessica Isserow, Justis Koon, Matthias Jenny, Arnon Levy, Agustín Rayo, Julia Markovitz, Daniel Muñoz, Matthew Scarfone, Russ Shafer-Landau, Jack Spencer, Bob Stalnaker, Katia Vavova, Preston Werner, Quinn White, and Steve Yablo. Finally, thanks to Jennie Kim, Tony Joe, Mimi Zander, Elaine Joseph, and Edwin Mateo for invaluable support these last few months, and thanks to my family – my brother Nayib Saillant and my parents Tania Montisano and Marco Saillant – for their unwavering confidence in me.

This work has also benefited from presentation at several places. Thanks to audiences at Aarhus University, Autonomous University of Madrid, Copenhagen University, Gdansk University, Ghent University, MIT, Syracuse University, Tufts University, University of Campinas, and VU University-Amsterdam. I am especially grateful to the folks at the Center for Moral and Political Philosophy at the Hebrew University of Jerusalem for inviting me to participate in their PhD Summer Workshop in Biology and Ethics and for their generosity and hospitality throughout.

CHAPTER 1

The Epistemic Upshot of Adaptationist Explanation

Philosophers, most recently metaethicists, have made multifarious things of the evolutionary development of our species. Some consider evolutionary science to be an epistemic godsend, which somehow or other provides us with the assurance that most of our beliefs are true or reasonably likely to be knowledge.[1] Others, by contrast, regard evolutionary considerations to be of little or no special relevance to our beliefs' epistemic standing, moral or otherwise.[2] Yet others feel, or fear, evolutionary explanation "debunks" moral belief: some argue it undermines the objectivity of morality;[3] others that it establishes our utter moral ignorance;[4] the rest seek a middle ground position between these extremes.[5] Thomas Nagel (2012) even suggests – indeed, insists – that the materialist neo-Darwinian conception of nature is, in some fundamental way, incomplete.

The debate over the epistemological implications of human evolution studies has tended to stray from the relevant scientific and empirical detail,[6] perhaps explaining a good deal of the divergence in opinion on these issues. But not all. In this chapter, I focus on understanding how learning about the evolution by natural selection of human cognition affects the epistemic justification of our actual beliefs.

---

[1] See, e.g., Quine (1975), Fodor (1983), Millikan (1984), Dennett (1987), and Dretske (1989).

[2] For the case against relevance in general see White (2010) and in the moral case Parfit (2011), Setiya (2012), and Vavova (2014).

[3] See, e.g., Ruse (1986), Street (2006; 2008), Kitcher (2011), and Bedke (2014).

[4] See, e.g., Joyce (2001; 2006), more weakly Fraser (2014), and with respect to only a proper subset of moral beliefs Singer (2005) and Greene (2008).

[5] See, e.g., Copp (2008), Enoch (2010), and Wielenberg (2010).

[6] See Fraser (2014), Kitcher (2016), and Deem (2016) for scientifically informed critiques of this unfortunate practice in the metaethical literature.

I argue that knowledge of the *adaptationist* (as opposed to spandrelist or exaptationist) explanation of a type of belief formation either defeats or bosters the justification of its output belief, or informs us of the fact that its output belief has never been justified. I argue more generally that when we learn about the evolutionary etiology of human belief formation we gain a form of *higher-order evidence*, roughly, evidence about the quality or reliability of human evidence. Along the way, I argue that a lack of specificity or clarity on the epistemological elements of the evolutionary debunking question has led to much of the confusion and disagreement found both among and across the different sides of the debate.

After foregrounding the pertinent epistemological and scientific concepts, I distinguish two types of adaptationist hypothesis according to whether the course of natural selection proceeds entirely without regard to the reliability of the target cognition, namely, *truth-orthogonal* and *truth-dependent* hypotheses (§1.1). I then argue that, if we learn a truth-dependent hypothesis is correct, we learn either that the target belief forming process is error-prone or that it is reliable, providing us with either a justification defeater or with further evidence that its output belief is justified (§1.2). Last I argue that, if we learn a truth-orthogonal hypothesis is correct, we learn that the target belief forming process is unreliable, informing us of the fact that its output belief has never been justified (§1.3). Throughout the chapter, as I develop the account, I apply it to various of the disputes among contributors to the evolutionary debunking debate in metaethics, clarifying some of its assumptions and implications. I conclude that much of the divergence in opinion is due to conflating two distinct types of adaptionist debunking explanations.

1.1 The Relevance, Target, Type, and Possible Upshots of Adaptationist Explanation

Evolutionary biology is relevant to epistemic evaluation because our means for representing the world, what I will call our *doxastic faculties*, are ultimately the product of biological evolution. Among known evolutionary mechanisms, natural selection stands out because differences in reliability can make for more or less prolific progenitors. On the assumption that these differences have a genetically heritable basis,[7] it follows that the adaptationist explanation of human cognition is a source of information on the origin of our reliability about everything. So, given that the origin of our reliability about a domain is

[7] This is no doubt a questionable assumption (Richardson, 2007; see also Sterelny, 2003), but it is widely held among contributors to the evolutionary debunking debate and, in any case, in exploring the relation between epistemic evaluation and adaptationist explanation, it must be taken for granted (Kahane, 2011).

relevant to the epistemic status of our beliefs about that domain, the adaptationist explanation of our doxastic faculties provides us with information about our beliefs' epistemic standing.

The precise significance of natural selection depends on (i) the way information about reliability or unreliability is relevant to the epistemic assessment of our beliefs, (ii) the relation between doxastic faculties and their genetic correlates, and (iii) whether and how differences in reliability have been selectively significant. In what follows, I treat each of these matters in this order.

1.1.1 Evolutionary arguments

As I will understand them, *evolutionary arguments* recruit adaptationist explanations of a doxastic faculty either to undermine or to further support its output belief.[8] Though other sorts of evolutionary explanation may in principle be recruited for these purposes,[9] for terminological continuity I follow the practice of referring to epistemological arguments based in adaptationist considerations with the broader term.[10] I call evolutionary arguments with positive import *buttressing* and those with negative import *debunking*. To buttress or debunk, evolutionary arguments disclose information about the reliability of the belief formation of the target beliefs or its enabling doxastic faculty or faculties. If jealousy makes me excessively suspicious of my partner, then my overblown suspicions may be discredited on account of their source in unwarranted fears of infidelity. If my obsessive-compulsive disorder makes me extraordinarily meticulous in the construction of my mathematical proofs, then my belief in the theorems may be further bolstered on account of its source in unreasonable fears of refutation. It is natural to think that (learning about) the reliability of beliefs' formation has repercussions for their epistemic standing.

---

[8] For simplicity, I assume the explanations recruited are known, not merely epistemically justified. I leave for another occasion discussion of the epistemic upshot of mere justified belief in these adaptationist explanations.

[9] In Chapter 3, for instance, I sketch a hypothesis of the cultural evolution of morality to undermine our confidence in the existence of moral knowledge.

[10] Others, e.g., Joyce (2016) and Clarke-Doane (2015), prefer the term *genealogical* to *evolutionary*. This move obscures the fact that the explanations recruited appeal to ultimate, as opposed to relatively distant, explanatory factors. As we will see in the text, this distinction makes an epistemic difference.

In general, there are three possible upshots of adaptationist explanation. Either the target faculty evolved to be reliable, unreliable, or anti-reliable. As it is to be understood in what follows, reliability is not a function of actual track record but has instead counterfactual force:

> *Reliability*: A doxastic faculty is *reliable* just in case it tends to produce true beliefs in the sorts of situations in which it normally functions.

On the pertinent conception of tendency, if normal conditions obtain (whatever those happen to be), then it is likelier to produce true belief than false if reliable and false belief rather than true if anti-reliable.

> *Anti-reliability*: A doxastic faculty is *anti-reliable* just in case it tends to produce false beliefs in the sorts of situations in which it normally functions.

A doxastic faculty's *degree of truth-conduciveness* is therefore the extent to which it tends to produce true belief in the admissible circumstances. A maximally truth-conducive faculty is perfectly reliable and a minimally truth-conducive faculty is perfectly anti-reliable. A faculty is neither reliable nor anti-reliable, and thus *unreliable* if and only if it lacks both the tendency to produce true belief *and* the tendency to produce false belief.[11]

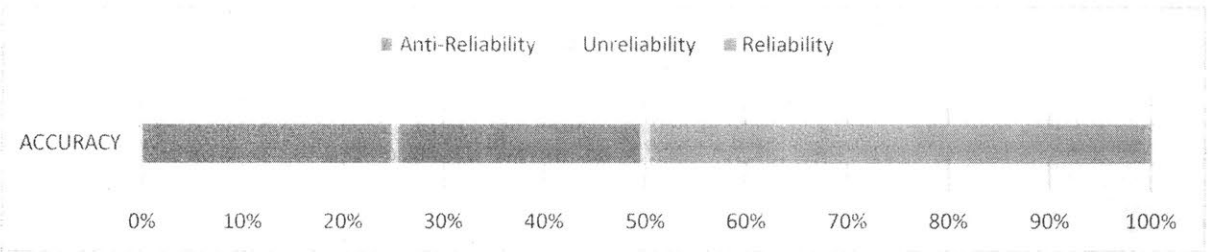

Fig. 1. Degrees of truth-conduciveness.

---

[11] This conception of reliability will suffice, but it is worth noting that it cannot capture interpersonal differences in beliefs' verisimilitude, i.e., their closeness to the truth, which may have also been selectively significant. The amended conception might also be able to accommodate truth-related variation across beliefs targeting necessities, also a limitation of the present framework. Belief in necessary or impossible truths may vary across individuals in a way that evinces more, or less, deeply flawed thinking.

1.1.2 Genes and doxastic faculties

It is important to understand that the relation between doxastic faculties and their genetic correlates because the mode of transmission in evolution by natural selection is genetic (Fisher, 1930). Thus, the effect of natural selection on the evolution of a trait, a doxastic faculty included, hinges on whether individual differences in its genetic correlates cause different rates of reproduction (Godfrey-Smith, 2007). In this section, I say a word on both the nature of the relation and its relata.

First off, about the relata, a *doxastic faculty* is the set of perceptual and/or cognitive abilities that support the formation, maintenance, and cessation of belief in propositions with a common subject-matter.[12] When your eyes dart around, capturing the color, shape, and size of surrounding objects, you exercise perceptual abilities that inform belief revision and so belief maintenance and cessation. When a twig nearby snaps and you automatically infer the presence of an agent (Barrett, 2000), you exercise cognitive abilities that result in belief formation. A faculty's *genetic correlates* consist in the genome regions responsible for the development of the cognitive and perceptual mechanisms that enable the formation, maintenance, and cessation of the faculty's output beliefs, where these mechanisms are neurally realized. The fusiform face area, a part of the human visual system, is specialized for facial recognition (Liu, Harris, & Kanwisher, 2010), enabling belief formation about face parts and face configurations. Genetic correlates underlie and regulate the development of cognitive and perceptual mechanisms and these enable, in turn, the exercise of perceptual and cognitive abilities.

Now, about their relation, it is causal, not informational (Godfrey-Smith, 2000; 2007). Doxastic faculties are not genetically encrypted. That would require, absurdly, the encoding of acquired characteristics, namely their constituent abilities, into the very molecular structure of DNA (Dawkins, 1982). Rather, a faculty's genetic correlates supply us with the potential to develop the faculty, a potential which is made actual if the appropriate environmental conditions obtain. The face recognition mechanism can only develop if the organism's senses are stimulated enough within a critical period to be able to detect face parts. Genetic correlates contribute to the suite of causal factors that determine the

[12] When relevant, I explain how faculties and their domains are to be individuated in the text.

course of a faculty's psychological development. For simplicity, I shall often speak of doxastic faculties as naturally selected, I mean to be understood as saying that their genetic basis, which underlies and regulates their development, is the result of protracted patterns of selection.

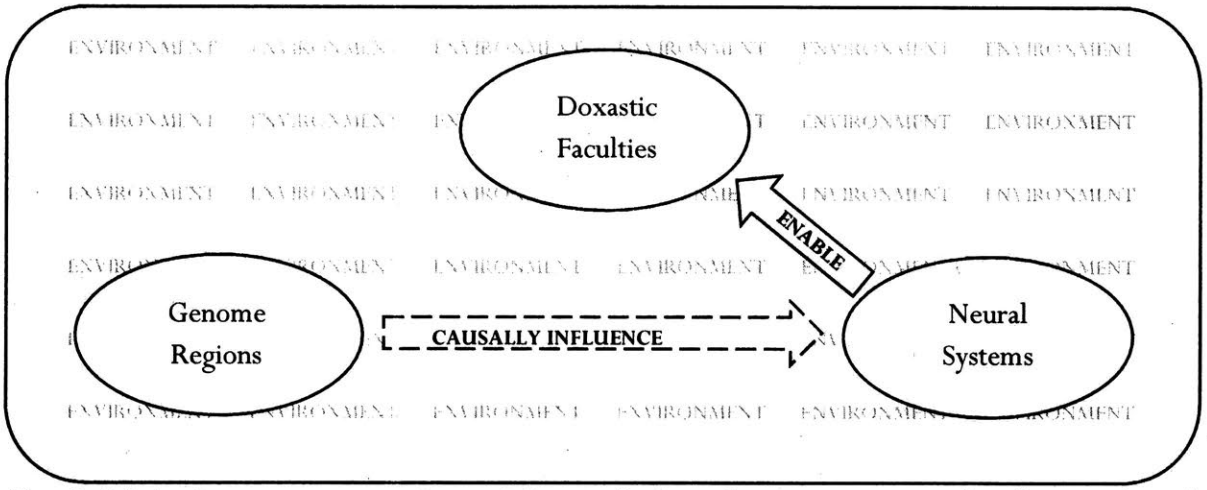

Fig. 2. Relation between genome regions and doxastic faculties. Genome regions causally influence, alongside environmental factors, the development of the neural structures that enable a doxastic faculty's constituent cognitive and perceptual abilities.

Let us now turn to the question of natural selection's influence on our beliefs. Since it is genetically transmittable entities that are naturally selected, and doxastic faculties aren't so transmittable, natural selection influences our beliefs by fixing the factors that help genetically to determine the present form of our doxastic faculties. In particular, by fixing the contents of our innate endowment, natural selection determines what faculties we have the *capacity* to develop. It therefore sets the upper limit of our epistemic potential because we can only believe the propositions for which we may develop the requisite faculty and, more importantly, as we'll see in §3, even if we do have the requisite faculty we may lack the *capacity* to acquire knowledge or even justified belief on its basis.

Lastly, a faculty's genetic correlates may vary across individuals and, consequently, in reliability. Differences in the genetic factors that contribute to the actual development of our faculties may interact with the relevantly same environmental factors in ways that lead either to more, or less, truth-conducive faculties. Genetic variation may therefore correlate, positively or negatively, with individual differences in reliability. We can think of the natural selection of our doxastic faculties as favoring degrees of truth-conduciveness greater than, or lesser than, that of our ancestors' contemporaries. I call the former

selection pressure *positive drag* and the latter *negative drag*. In the following section, I draw the more general distinction between selection that involves the target truths and selection that does not.

1.1.3 Truth-dependent vs. truth-orthogonal selection

Throughout the adaptive evolution of a doxastic faculty, differences in reliability may have been selectively significant, but it's also possible that they were selectively *in*significant. If significant, then different degrees of truth-conduciveness caused different rates of reproduction and over time a particular degree came to predominate (*Truth-Dependent Selection*).[13] Positive and negative drag are subtypes of truth-dependent selection. However, if differences in truth-conduciveness were selectively insignificant, then no degree of truth-conduciveness is favored by natural selection and so the faculty's adaptation proceeded entirely orthogonal to the truth of its output belief (*Truth-Orthogonal Selection*).

According to truth-dependent hypotheses, our ancestors outreproduced their contemporaries because their faculties were more, or less, truth-conducive than theirs. Their beliefs, because of their truth-value or their proximity to the truth, led to a greater amount of adaptive behavior, or behavior more adaptive. In the truth-dependent evolution of a faculty, the accurate, approximate, or inaccurate representation of facts is reproductively profitable. One would expect many perceptual doxastic faculties, for instance, to have evolved by positive drag, since being more reliable about one's immediate environment than one's peers, predators, or prey clearly confers a reproductive advantage.[14] Moreover, since risk may be averted as a result of systematic error-proneness, one can also easily imagine evolution by negative drag. The so-called *sexual overperception bias*, for instance, may have led our forefathers to outreproduce their competitors because a tendency to attribute sexual interest and intent to women based solely on friendliness may have made overlooking sexually receptive females less likely (Haselton, 2003). In truth-dependent selection, belief helps generate adaptive behavior because of its correlation with the truth, positive or negative.

---

[13] This gloss assumes for simplicity that the type of selection in play is directional. Directional selection occurs when natural selection favors one extreme of continuous variation, so in this case either extremely high or extremely low degrees of truth-conduciveness.

[14] But see Mark, Marion, and Hoffman (2010) for an argument that veridical perception can be driven to extinction by non-veridical mind-world interfaces.

According to truth-orthogonal hypotheses, our ancestors reproduced more prolifically than their contemporaries because of differences in truth-irrelevant features of the output belief. In the truth-orthogonal evolution of a faculty, the truth-value of the propositions believed is biologically irrelevant. Johnson (2015) argues, for example, that the supernatural doxastic faculty first emerged as a by-product of distinct doxastic phenomena and over subsequent generations was adapted to enable cooperation among non-kin. He hypothesizes that belief in the presence of moralizing deities increases prosocial behavior by creating the impression of being watched when alone and by inducing the fear of supernatural punishment for non-cooperation. The so-called *supernatural punishment hypothesis* is truth-orthogonal because the existence of moralizing supernatural agents isn't needed for belief in their presence to facilitate and maintain cooperative behavior and is therefore unnecessary to the explanation of supernatural belief's presence and prevalence. In truth-orthogonal selection, belief helps generate adaptive behavior irrespective of its correlation with the truth.

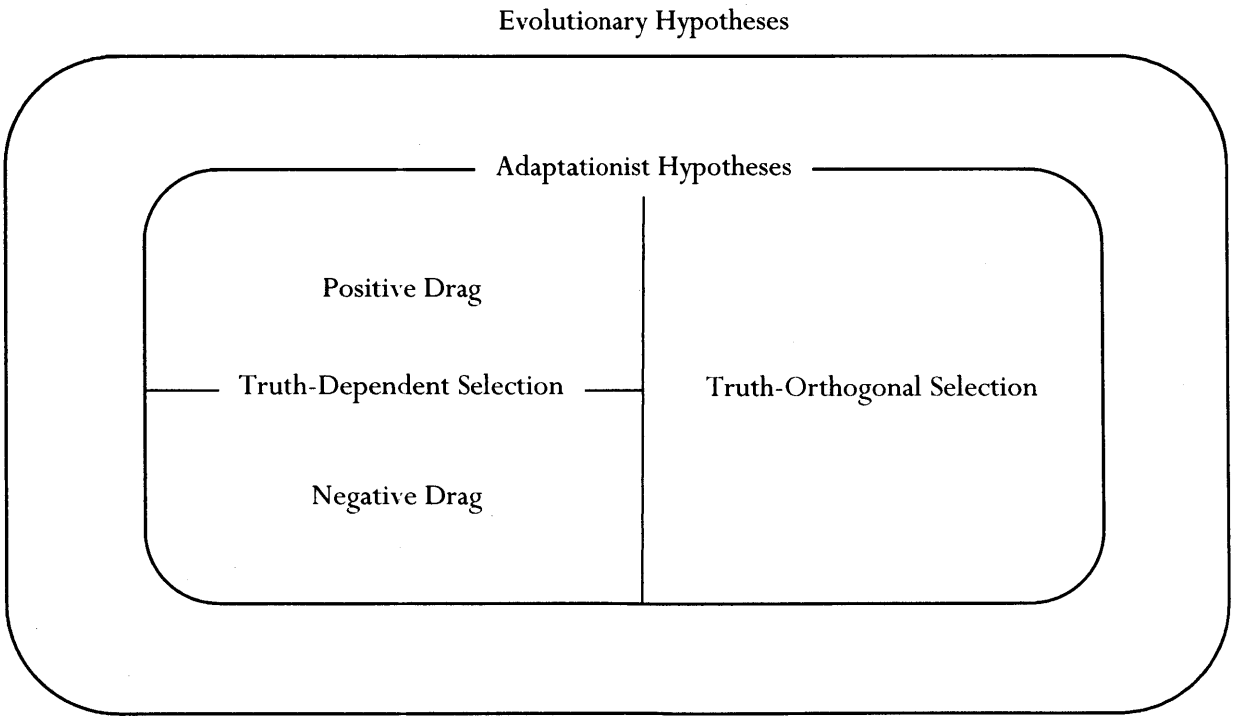

Fig. 3. Space of evolutionary hypotheses about human cognition. The adaptationist hypotheses are (i) a proper subset of evolutionary hypotheses, (ii) appeal either to truth-dependent or truth-orthogonal selection, and (iii) if truth-dependent appeal to either positive or negative drag.

If our doxastic faculties are adaptations, then they evolved either truth-dependently or truth-orthogonally. As we will see, much of the diversity in the evolutionary debunking literature results from conflating truth-orthogonal selection with negative drag. I first turn to examine the upshot of truth-dependent adaptationist explanation.

1.2 The Epistemic Upshot of Truth-Dependent Explanation

In truth-dependent explanation, we were naturally selected to have the potential to become truth-conducive to some degree about certain domains. Selection on our lineage explains why we have the genetic correlates of truth-conducive versions of a faculty. However, in general, it takes very many rounds of selection for a genetic variant to predominate. Therefore, the cogency of arguments from truth-dependent explanation, what I will call *drag arguments*, depends on both the strength and the duration of the hypothesized selection pressure. For one's confidence that one has the correlates of the target faculty directly depends on how many rounds of selection have transpired and on how reproductively advantageous the favored variant is in comparison to the rest. For simplicity, I will assume that the selection pressures referenced in drag arguments have been in play long enough for the favored variant to spread across the population.[15] I begin with positive drag arguments.

1.2.1 Positive drag arguments

Philosophers have tended to focus on the epistemically negative side of adaptationist explanation. Though this is natural given that that's where the skeptical worries arise, failing to keep in mind the positive counterpart has led many to adopt, at least implicitly, an overly narrow conception of selectionist considerations and in the end, I think, helped many run truth-orthogonal selection and negative drag together (Clarke-Doane (2012) is a case in point; see §4.5). In this section, I explain positive drag to underline the diversity of selectionist considerations and argue for an account of its epistemic upshot for use as a foil helpful in thinking about the upshot of the rest.

---

[15] Recall also that for simplicity we are assuming that the type of selection at work is directional (see fn. 13). So, the favored variant is to be taken to be as close to the extremes as biologically possible.

Under positive drag, a faculty's truth-conduciveness ratchets up across the generations. In this scenario, part of what explains why each of us exists is that our forebears outperformed their contemporaries in the acquisition of true beliefs about a certain domain. If any of the doxastic faculties of a species evolved by positive drag, then the observation that we are members of this species is excellent evidence that every one of us has a reliable version of the faculty, since by hypothesis our parents' reproductive achievement is in part due to their possession of more truth-conducive versions themselves as is that of their parents and so on up the genealogical tree. The underlying principle

> POSITIVE DRAG BUTTRESSING: If we learn our existence is partly due to a faculty's greater truth-conduciveness in our ancestors, we have sufficient evidence to believe that it is reliable in us.

I write "our existence" to sidestep the question of whether natural selection explains individual traits or just the distribution of traits across a population, a matter of controversy in the philosophy of biology.[16] Both sides agree, however, that our existence is of course explained, since our ancestors' reproductive success explains why each of us exists (rather than not exist) simply due to their place in our respective lineages. Given this and that *ex hypothesi* the target faculty is heritable and its greater truth-conducive operation has been adaptive for long enough (as per the section-opening simplification), the principle states that that suffices to believe with justification that it's reliable in us.

Take our perceptual doxastic faculties, for example. If our existence is partly due to our ancestors' accurate representation of their immediate environment, then it's highly likely that our endowment includes the capacities needed to develop reliable perceptual mechanisms in environments that relevantly resemble the natural habitat of our ancestors. Since environments didn't change radically enough from one generation to the next to render the following generation's endowment non-adaptive (otherwise either we wouldn't exist or the apocalypse is nigh), our present environment must resemble that of our ancestors as needed for the development of perceptual reliability. More generally, on a positive drag hypothesis it is highly likely that the target faculty is reliable in us because we can assume

---

[16] Mogensen (2015) argues that debunking arguments assume selection explains individuals' traits, a mistake that, he thinks, should cause debunkers to "give up the emphasis on selection" (p. 17). While he may be right about past debunkers assuming this, as I will explain in the text, to debunk (or buttress) selection need only explain our existence.

that the environment didn't change sufficiently between generations to render its normal development impossible or its truth-conducive operation non-adaptive.

A positive drag explanation, moreover, provides *sufficient* evidence because its truth implies that the faculty evolved to be reliable. Since we've *learned* that it evolved by positive drag, the evolutionary change supposed to have transpired just is the development of its tendency to form true beliefs, that is, the evolution of its reliability, since it originally was neither reliable nor anti-reliable and, as I stipulated for simplicity at the outset, high degrees of truth-conduciveness have been favored for long enough to spread across the population. The knowledge that a faculty evolved by positive drag provides us with evidence sufficient for believing, with adequate justification, that the evolved faculty is reliable in us.

While sufficient for justified belief, the evidence is not conclusive. In some cases, it may even be misleading. If my faculty happens to be anti-reliable, a positive drag argument would still supply me with evidence for its reliability because it continues to be the case that the correct explanation of the fact that I exist appeals to my ancestors' possession of the more truth-conducive versions of the faculty. It just so happens that in my case other factors – perhaps environmental, perhaps not – managed to render it anti-reliable. Positive drag explanations provide us with sufficient evidence for justified belief in our faculty's reliability, but only in the absence of defeating considerations.

Let's suppose however that the evolutionary evidence for reliability is not undercut for you and that on its basis (and on that basis alone) you believe that your version of the faculty is reliable. Should that impact the epistemic standing of the beliefs you hold on the faculty's basis? Consider:

> **Reliability Pill:** Unbeknownst to you, you are administered a drug that greatly increases your doxastic faculty's truth-conduciveness. Once the drug is in effect, you are given a task that exercises it and as a result you form the belief that *p*. You are then informed of your predicament.

It's clear that this would be a piece of good news about your belief that *p* because it would tell you that it's likelier to be true than it would otherwise have been. It should also be clear that whatever benefits your belief reaps from the increase in reliability (be it in the form of greater sensitivity or greater safety from error or what have you) would be bestowed upon the belief at formation because the report is merely *about* the facts that set the belief's epistemic standing. Upon hearing the news, your belief's

epistemic standing as knowledge or justified is unchanged and at most it's only your (higher-order) belief about its standing that would merit revision. Indeed, the news that your belief was formed under the influence of the drug doesn't even tell you that it's an item of knowledge, for at most it just tells you that it's more justified than it would have been otherwise, not that it's sufficiently justified to be knowledge.

But what if you're *not* a reliabilist about justification? Why would it tell you that the belief is *at all* justified in the first place? Recall the purpose of considering **Reliability Pill**: to help us understand how learning that a faculty evolved by positive drag, and thus that it's reliable in us, should impact the output's standing as justified or knowledge. This exercise would be pointless unless we assume that the possibility in which you are administered the drug is as close as possible to the way things actually are. The reason why an increase in reliability signals an increase in justification, even if you aren't a reliabilist, is that *in the real world* one becomes more reliable on account of becoming better at gathering or evaluating evidence. In the real world, that doesn't just magically or inexplicably happen. Put differently, the point is that we can legitimately conclude that there is an increase in justification from an increase in reliability without being a reliabilist because in the real world the best, if not only, explanation of such an increase is that one became better at forming beliefs with epistemic justification.

So far, the lesson drawn from **Reliability Pill** is that news of a faculty's greater reliability doesn't *improve* the output beliefs' standing as knowledge or justified and that it doesn't tell us whether the output beliefs amount to knowledge. However, it does tell us that the beliefs are more justified than they would otherwise be, since at most otherwise they would be completely unjustified and thus in actuality they must be more justified than that. Learning this provides us with the confidence to hold onto the output beliefs in the face of *justification* defeat. The underlying principle is

> REAFFIRMING BUTTRESSING: If you have sufficient evidence to believe that your doxastic faculty is reliable, then you have sufficient epistemic reason to hold onto its output beliefs.[17]

---

[17] From antecedent to consequent I move to the notion of epistemic reason because the evidence concerns the higher-order matter that are the circumstances of belief formation and the recommended conservativism concerns the lower-order beliefs. Since the lower-order beliefs aren't generally about the circumstances of their formation, it would be inappropriate, strictly speaking, to construe the evidence about belief formation as being evidence in favor of the

On encountering defeasible evidence against the truth of an output belief, the principle tells us to continue believing that *p if we still have sufficient evidence to believe that the belief was formed reliably*. In **Reliability Pill**, if an epistemic inferior tells you that not-*p*, you still have sufficient reason to continue believing that *p* because your belief that your belief that *p* was formed reliably continues to enjoy the support derived from news of your predicament. In the case of positive drag, then, since we have sufficient evidence to believe that our faculty is reliable, that is, likelier to produce true belief than false under normal conditions, we should continue to hold onto the lower-order beliefs unless we have reason to think that conditions were abnormal or that there is evidence sufficient to defeat one's justification for the belief that one's version of the faculty was, on that occasion, truth-conducive.

Lastly, it is important to note that the causal factors at work in cases of positive drag are ultimate, not proximate. Positive drag explains why we have the genetic correlates we do and thus why we are prepared to develop (in an environment such as ours) a reliable, rather than an anti-reliable or unreliable, version of the faculty. It explains the *origin* of our reliability about the faculty's domain. Drug action, for instance, explains a temporary surge in reliability, but critically, in the real world, the success of such a drug is contingent on the existence of the *capacity* to be reliable. Proximate explanation may at most indicate one's greater reliability than before or than those unaffected by the proximate factor. By contrast, positive drag explanation tells us that the species-typical faculty is reliable. Evolutionary boosters may recruit adaptationist explanations to support the reliability of doxastic faculties and in consequence the justification of its output belief.

1.2.2 Negative drag arguments

Under negative drag, a faculty's truth-conduciveness ratchets down across the generations. In this scenario, our existence is contingent on our forebears' misapprehensions. We exist partly because of our parents' reproductively advantageous misconceptions and illusions, and they due to that of theirs and so on for the duration of negative drag. The observation that we are members of a species with a

---

lower-order beliefs. Rather, this higher-order evidence provides us with an epistemic (as opposed to pragmatic) reason for holding onto the beliefs.

history of negative drag provides us with excellent evidence that we share in our lineage's reproductively advantageous biases. The underlying principle is

> NEGATIVE DRAG DEBUNKING: If we learn our existence is partly due to a faculty's lesser truth-conduciveness in our ancestors, we have sufficient evidence to believe that it is error-prone in us

The same argument given for POSITIVE DRAG BUTTRESSING can be recruited, *mutatis mutandis*, to support this principle. Again, the environment cannot have changed sufficiently across generations to render error-proneness non-adaptive or its development impossible and, again, since at the outset a faculty is unreliable the correctness of such an explanation entails the evolutionary development of its error-proneness. The parallel lessons also apply. While sufficient for justified belief in the absence of defeating considerations, the evidence isn't conclusive and may even be misleading.

We have sufficient evidence for error-proneness rather than full-blown anti-reliability, for, as McKay and Dennett (2009) note, systematic bias is only adaptive against a background of true belief. Recall the male tendency to over-attribute sexual interest to women. According to the male sexual overperception hypothesis, this bias in social cognition helped our forefathers out-sire their competitors because it made missed mating opportunities less likely. But for the bias to actually make them more efficient *progenitors*, ancestral men would also need to believe *truly* that the targets of the attribution are (a) *living* organisms, (b) of their *own* species, (c) of the *opposite* sex, (d) of *child-bearing* age (etc.). They would need to first get a great deal right about their social world. Indeed, the bias is to mistake *friendliness* for sexual interest (Abbey, 1982). Negative drag arguments cannot present domain-wide skeptical threats because under negative drag a faculty can only develop bias, a warp in its otherwise truth-conducive operation.

Further, we can learn of an adaptive bias's existence without any evolutionary theorizing because psychological investigation can uncover it all by itself. Since adaptive biases don't affect entire domains of belief or forms of reasoning, we have a background of reliability against which we can check for their distorting, error-inducing influences (that's why psychologists are able to discover the biases in the first place). At any rate, the evolutionary theorizing is unnecessary because the bias's status as an adaptation is epistemically irrelevant. *How* we learn of the bias is immaterial to its epistemic significance because it

is its *existence* that casts doubt on the truth of the biased output. Since psychological experimentation would do just as well (and adaptationist explanation is so epistemically demanding (Gould & Lewontin, 1979)) it is, to put it mildly, dialectically unwise to rely on evolutionary rather than psychological explanation. In contrast to their buttressing counterparts, negative drag arguments are therefore *not* indispensably evolutionary, since psychological explanation may be recruited to justify the same conclusion.

I have already argued enough to conclude that if Street and Joyce's arguments were intended to be negative drag arguments, neither can work to establish their conclusions because both Street and Joyce believe that *entire* domains of belief are to be debunked. But since negative drag can only motivate the *partial* revision of a faculty's output belief, and the difference in subject matter between beliefs about different domains would require the exercise of distinct sets of abilities, proper subsets of output belief may not be about different domains because – since a faculty just is a set of perceptual or cognitive abilities – that would only be possible if *distinct* faculties were in play. So, since the selectionist explanation of relevance just concerns the one faculty whose evolution is to be explained, a negative drag argument cannot motivate domain-wide skeptical implications and thus neither Joyce's nor Street's argument could work as an argument from negative drag.

However, those that aim to debunk just a subset of moral belief – *modest debunkers*, as I will call them – must, at least implicitly, assume the negative drag of the moral doxastic faculty, and there are very many of these philosophers, it seems. In "Evolutionary Debunking Arguments," Guy Kahane observes that:

> It is now common to think of nature as a *distorting* influence on our evaluative beliefs [...]. And implicit or explicit debunking arguments that rely on this assumption are widely used in contemporary evaluative ethics. (2011, 109; emphasis in original)

If he is right (and he does mention five different examples),[18] these philosophers are relying on something like:

---

[18] Parfit (1984), Singer (2005), Crisp (2006), Greene (2008) and Huemer (2008), specifically.

UNDERMINING DEBUNKING: If you have sufficient evidence to believe that your doxastic faculty is error-prone within certain subdomains, then you have some epistemic reason to abandon its error-laden output.[19]

I write "error-laden" rather than "erroneous" to emphasize the gradability of the error. The sexual overperception bias, after all, involves only a very minor piece of misrepresentation compared to everything else one must get right. Furthermore, the error-laden beliefs are presumably *un*justified (and thus not knowledge), since in the real world such a bias in the formation of these beliefs signals a misappraisal of the evidentially relevant facts. A man's erroneous belief that the amiable woman in front of him is sexually attracted to him is unjustified because contrary to what he thinks there just is no evidence of sexual interest.

Modest debunkers face several serious problems, the most important of which is figuring out which beliefs are error-laden. To use UNDERMINING DEBUNKING one must identify which subset of the moral faculty's output beliefs is error-laden, and that depends, quite problematically, on the details of the correct selectionist explanation. The problem here isn't just that it is unclear whether we know enough about its evolutionary history to tease the two sets of beliefs apart. The problem is that we know enough to strongly suspect we will never be able to do this. Indeed, Lewontin (1998) argues that we will never be able to explain the origin of cognition, let alone the human moral kind, because we lack, among very many other things, the fossil record to reconstruct its evolution.

Moreover, assuming for argument's sake that the dearth of fossil evidence isn't prohibitive, there is widespread agreement among evolutionary scientists that if our moral doxastic faculty did evolve adaptively it involved *gene-culture co-evolution*. This would mean that the cultural transmission of moral opinion influenced the genetic transmission of the moral faculty's correlates (and vice versa) over its evolution, and as Kitcher (2016) notes cultural selection may favor traits natural selection would have eliminated. So acculturation may either counteract or compound distortion, which makes the task of isolating the error-laden beliefs practically impossible. So, when Peter Singer recommends that we

---

[19] For a defense of this kind of principle, see Christensen (2010) and Horowitz (2013).

"attempt the ambitious task of separating those moral judgments that we owe to our evolutionary and cultural history, from those that have a rational basis [...] for it is the only way of avoiding moral skepticism" (2005, p. 351), he sorely underestimates the scale and difficulty of the task. To suppose it practicable, as all modest debunkers must, is naively optimistic.

Lastly, Kahane (2011) suggests that modest debunkers face an overgeneralization problem. He wonders whether "the [evolutionary debunking argument] can be stopped from covering the whole of the evaluative—from supporting, not utilitarianism or even rational egoism, but global evaluative skepticism" (p. 114). If it is a negative drag argument (and Kahane takes all evolutionary debunking arguments to be negative drag arguments), the answer is YES, because there would be a matter of fact concerning which of its output are error-laden and which are not. So, quite to the contrary, the explanation (if we could learn it) would tell us which are unbiased and thus by REAFFIRMING BUTTRESSING would serve to *protect* their justification from defeat, let alone fail to debunk them. But, and perhaps this is what Kahane was sensing, modest debunkers do face the problem of justifying the adoption of a negative drag hypothesis over the truth-orthogonal alternative and it is unclear what sort of evidence or rationale they could use to favor a negative drag hypothesis over a truth-orthogonal hypothesis with just as much empirical support.

1.2.3 Summary

In brief, there are two types of arguments from truth-dependent explanation, one buttressing and the other debunking. Under ideal conditions, the former tells us that the faculty in question is reliable and the latter that a proper subset of the faculty's output belief is error-laden. But conditions are not ideal, in two important respects.

First off, the simplifying assumption that the relevant selection pressures have been around long enough for the favored variant of the faculty to spread across the population would require a great deal of evidence for its justification. Since a member of the population may have a version of the faculty without its being the favored variant, to justify the assumption it wouldn't be enough to note that all members of the population have the relevant beliefs. Further, since we cannot determine how much the favored variant has spread, we cannot determine the extent to which a drag argument justifies its

conclusion because we wouldn't know how likely one is to possess the hypothesized degree of truth-conduciveness. Second, just as acculturation may correct distortion, it may also foment it and via gene-culture co-evolution turn the course of evolution from positive to negative drag. So even if one subtype of truth-dependent explanation is known to be correct, it needn't follow that we are presently such that the faculty operates reliably rather than fallibly.

Under realistic conditions, even if we were to learn a truth-dependent explanation is correct of a certain faculty of ours, it would remain quite obscure how much credence one should invest in the conclusion of the drag argument in question. In practice, for the time being at least, truth-dependent explanation would be useless for the purposes of epistemic assessment.

1.3 The Epistemic Upshot of Truth-Orthogonal Explanation

Truth-orthogonal hypotheses appeal to selection pressures on cognition that proceed in complete indifference to the target truths. A doxastic faculty helps its bearer outreproduce competitors by forming beliefs that aid survival and reproduction irrespective of their truth-value. So, degree of truth-conduciveness has no explanatory role in a faculty's evolution because differing degrees of truth-conduciveness *do not* cause different rates of reproduction. Rather, it is individual differences in a faculty's truth-indifferent operation that has such an effect. In truth-orthogonal explanation, our ancestors outreproduced their contemporaries because their faculties outmaneuvered theirs neither *because* nor *despite* their relative truth-conduciveness. The entirety of a doxastic faculty's selective significance lies in the production of behavior and none of it in its getting things right or wrong because how matters *actually* stand with respect to their subject matter is biologically irrelevant.

In this section, I argue that arguments from truth-orthogonal explanation, *drift* arguments as I will call them, inform us of the fact that the output beliefs cannot amount to knowledge or become epistemically justified.

1.3.1 Truth-orthogonal debunking

A truth-orthogonal explanation informs us of the fact that a faculty's biofunction, its evolutionary *raison d'être*, is unrelated to the acquisition of true belief because its output beliefs' status as true rather than false, or false rather than true, is selectively insignificant. The relevant principle is:

> TRUTH-ORTHOGONAL DEBUNKING: If we learn our existence *isn't* due to a faculty's truth-conducive operation in our ancestors, we have sufficient evidence to know that faculty was not adapted to be either reliable or error-prone.

If we discover that a doxastic faculty evolved truth-orthogonally, we learn its genetic correlates were naturally selected regardless of its degree of truth-conduciveness. Outlandish possibilities aside,[20] the observation that we exist is conclusive evidence that the faculty is unreliable on the assumption that there is no other non-causal connection between human moral psychology and moral facts.[21]

But, I should note, even setting aside the possibility of non-causal psycho-moral connections, it continues to be metaphysically possible that the resulting faculty is reliable. For instance, purely coincidentally, across the entire evolution of a faculty, humans may have occupied environments in which the output belief happens to be true even though its truth-value plays no role in explaining its adaptiveness. For example, on the assumption of utilitarianism, an evolutionary account on which our existence is partly due to our ancestors' utilitarian beliefs may be truth-orthogonal, but leave us nonetheless with reliable moral faculties. While metaphysically possible, if it occurred, it would be a biological miracle.

---

[20] It is compatible with a faculty's truth-orthogonal evolution that if the faculty's subject matter exists, some extraterrestrial alien with a reliable counterpart of the faculty, and the power to repair ours, has – unbeknownst to us – made us reliable (or systematically error-prone) about the domain in question.

[21] As I discuss in Chapter 2, affirming the existence of such a psycho-moral nexus does not help forestall the evolutionary challenge as I understand it. Even moral subjectivists face an epistemological challenge from evolutionary considerations.

Consider that for something like that to happen the course of truth-orthogonal selection must simulate a course of positive drag.[22] Not only must genetic material capable of being worked into the correlates of a reliable version make it into the gene pool, it must also be the case that genetic variants associated with relatively greater degrees of truth-conduciveness happen also to luckily be associated with the selectively advantageous features of the faculty. Secondly, to assume this occurred would be unmotivated twice over and in two ways, since both the supposed presence of the right genetic material and its supposed association with the selectively significant features lack, from a biological standpoint, empirical *and* theoretical motivation. Indeed, to *become* motivated the supposition that the right genetic material is available requires the reality of the domain in question.

But here's the real kicker. Let's assume the mind-independent reality of the moral domain. If you are willing to ignore the possibility of supernatural intervention, then it follows that at any point in the course of the faculty's evolution, over a period of at the very least tens of thousands of years, what was adaptive to believe about moral matters might have easily diverged from the reality of the matter because *ex hypothesi* what happened to be adaptive in that environment wasn't so *because* of its truth. For had the two been incapable of coming apart then its descent would instead be truth-dependent. If the truth and the adaptiveness of moral belief ever lined up in the truth-orthogonal case, it would have to be accidental, since truth and adaptiveness would not be explanatorily connected (causally or otherwise). For the faculty to turn out reliable, then, this coincidence would therefore have to be maintained, as well as (from an epistemic standpoint) accidentally improved upon, over countless generations and across huge expanses of land. It would be a Rube Goldberg machine like no other, not just in magnitude or because it would be entirely accidental, but because the links in the chain reaction, while causal, wouldn't explain why it turned out to be reliable rather than an unreliable version of the faculty.

Indeed, this possibility would be so improbable given knowledge of its truth-orthogonal origins that taking it seriously enough to preclude knowledge of unreliability would have wide-ranging skeptical

---

[22] If you are wondering why there *must* be this kind of simulation and thus the absence of some extraordinarily convenient mutation, the reason is that selection would otherwise not *explain* the origin of the faculty, since it would instead be genetic mutation, a distinct mechanism, that explains it. It's through the multigenerational retention of myriad modifications that natural selection would account for its origin, which is what we are asked to supposed when told to assume the truth-orthogonal account is correct.

implications. If we were to learn that a faculty evolved truth-orthogonally but denied that this basis for belief in its unreliability suffices for knowledge, then we would lose our claim to knowledge of most, if not all, contingent truths. For if the vanishingly small probability of the truth-orthogonal evolution of a *reliable* faculty is enough to preclude knowledge of its unreliability (even if actual), then by parity of reasoning similarly improbable possibilities of falsehood would prevent knowledge of other contingent truths. In effect, we would be forced to embrace the conclusion of the lottery paradox because pretty much every contingent truth has a similarly, if not less, small probability of being false (Hawthorne, 2004). In the absence of suitably independent justification for reliability, if we are to abstain, rightly, from considering belief in unreliability on these grounds to amount to knowledge, we would have to accept broad-spectrum skepticism about contingent truths.

Suitably independent justification, moreover, will always be absent in the truth-orthogonal case. First off, the justification is only suitably independent if it doesn't ultimately advert to the content of the target beliefs because otherwise, in so doing, one would inevitably rely on the target faculty's reliability in one's attempt to justify it. Such an attempt would be viciously circular because the target of justification is precisely one's claim to reliability.

However, as Setiya (2012) observes, one still has the option of appealing to the target beliefs, and their contents, as "intermediate steps in an argument for reliability, not its ultimate ground" (p. 82). On his view about moral belief specifically, since one's source of justification for holding them are the nonmoral facts in virtue of which a moral proposition is, or would be, true, such an appeal doesn't necessarily lead to a vicious form of circularity, since neither one's beliefs, nor their contents (nor for that matter any other product of moral psychology), constitute one's evidence for their truth. Given that these nonmoral facts indeed justify one's moral beliefs, one may justify one's reliability about moral matters on the basis of one's (justified) self-ascription of these beliefs and one's belief in their content.

On Setiya's view, then, one's justification for moral beliefs, assuming for argument's sake that it exists, needn't be immediately viciously circular. However, the question remains whether it would be strong enough to withstand the discovery of a truth-orthogonal descent, since in such a case the nonmoral facts which we naturally take to be morally relevant would only be *correctly* so taken as a matter of adaptive happenstance. Indeed, from the evolutionary perspective, since we know the faculty to have evolved

truth-orthogonally, the more probable scenario is that we mistook morally irrelevant facts to be relevant to the truth of moral beliefs because there are very many more ways of getting things wrong than right, and there is no pressure to get them right.

Furthermore, this scenario seems to be the more promising empirical hypothesis due also to independent explanatory grounds. Specifically, it's prima facie puzzling exactly how facts about a domain manage to justify beliefs about facts of an entirely different kind, not to mention habitually so, let alone at the end of the day exclusively. The skeptical hypothesis would explain this with ease, since it would all be illusory. We would incorrectly, but in a psychologically compelling way, take nonmoral facts to be evidentially relevant. Even if by pure luck we took the right nonmoral facts to be relevant, these considerations would defeat this justification, since the alternative would be both better evidentially supported and abductively more secure. Given its defeat, the appeal to moral beliefs as an intermediate step would be illegitimate and so the argument for reliability would be deprived of its justificatory power. So, its use, while not *blatantly* circular, nevertheless is indirectly so in the end, since it helps itself to justification that is incompatible with the knowledge that our moral doxastic faculty evolved truth-orthogonally.

In sum, since there is no candidate source of justification left, there is no suitably independent justification in cases of truth-orthogonal explanation and thus if we are to avoid skepticism about contingent truths we must accept TRUTH-ORTHOGONAL DEBUNKING.

1.3.2 Drift arguments

So, what can we conclude from truth-orthogonal explanation? On the orthodox post-Gettier view, to amount to knowledge a belief must be based on grounds that are non-accidentally connected to its truth. Since unreliable faculties don't base their belief-forming activity on truth-relevant considerations, their output beliefs, if true, must be so accidentally and therefore cannot amount to knowledge. Drift arguments inform us of the fact that we are incapable of knowing anything on the basis of such a faculty because veritic luck infects every fiber of its being (cf. Pritchard 2007).

Indeed, the faculty cannot yield justified belief. Since the output beliefs are never based on truth-relevant considerations, they were never epistemically justified, because the considerations one is led to *view* as relevant to their truth are, as a matter of fact, immaterial to it. Consider:

> **Fortune Telling**: You're faced with a life-changing decision and, as you always do when uncertain about the future, you consult your personal fortune-telling device, your Magic 8-Ball. You ask whether *p*, give it a shake to the left, a shake to the right, and invest belief or disbelief in *p* with the indicated degree of confidence. You are then told that Magic 8-Balls are just novelty toys, devices which are for entertainment rather than truth-seeking.

Surely, to continue to have any measure of credence in either *p* or not-*p* would be unjustified because your belief that *p* is revealed to have never been based on truth-relevant grounds. Since the toy was never in contact with facts pertaining to your fortune, its deliverances were never a suitable basis for belief about your future or its goodness. And this is the case even if we add that you had *justifiably* thought it provided such a basis because, by the facts of the case, regardless of what you thought *it did not*, and simply thinking that it does (justifiably or not) does not change that fact.[23]

A truth-orthogonally constructed faculty relevantly resembles a Magic 8-Ball because both were engineered to fulfill non-epistemic purposes. Both were built in total disregard of their purported subject matter. They have an ulterior purpose, serving a function that lies beyond what seems evident in its overt behavior: in one case generating revenue and in the other offspring. The deliverances of both are groundless: one is just as unreliable an indicator of fortune as the other is of the facts of its putative domain. Belief held on account of the activity of one suffers from the same epistemic inadequacies as belief held in virtue of the activity of the other: neither issues deliverances that reflect the truth-relevant facts, for their ultimate aim is orthogonal to the truth. Neither is capable of generating knowledge or justified belief because their operation is *un*reliable and so non-truth-value-tropic.

---

[23] Of course, if you were justified in thinking it was a genuine fortune-telling device (say, a trusted informer played a prank on you, but – uncharacteristically – forgot to tell you it was a prank), you are certainly less rationally criticizable than had you thought this without any justification. But this is because of the way the belief about the 8-Ball's fortune-telling power was formed, not anything to do with the beliefs formed based on its utterly uninformative deliverances.

1.3.3 Individuating doxastic faculties

One might think that learning of the truth-orthogonal evolution of a faculty is not as devastating as it initially seems because while that faculty is epistemically bankrupt we may still have alternative means for belief about its domain. Many anti-debunkers, for example, claim that we can compensate for the "non-truth-tracking" evolution of morality through the use of domain-general reasoning or by relying on other doxastic faculties. This is a common suggestion. It rests, however, on a confusion concerning the individuation of doxastic faculties, for the conception of a faculty as a set of cognitive or perceptual abilities together with the fact that faculties must be individuated according to their adaptive function entails in the truth-orthogonal case the impossibility of self-correction. The following may seem a bit complicated but it is needed to appreciate this entailment.[24]

The first thing to note is that while faculties evolve either truth-dependently or truth-orthogonally the underlying cognitive and perceptual mechanisms may evolve truth-orthogonally with respect to one domain but not another. For example, according to the supernatural punishment hypothesis, the HADD, the hyperactive agent detection device (recall §1.1.2), is in part responsible for creating the impression of being watched when alone. If this suggestion is correct, then the HADD was adapted over subsequent generations to help yield supernatural belief while performing at the same time its original biofunction of alerting one to the presence of predators, prey, or peers. With respect to the latter subject matter, the HADD presumably evolved truth-dependently but with respect to supernatural agents it evolved truth-orthogonally. Since different faculties may share component mechanisms, the underlying wetware may have been subject, over evolutionary time, to multiple selection pressures at once.

The second thing to note is that a component mechanism may serve different faculties on account of enabling *different* abilities. For example, the eye is a component mechanism of the visual perceptual doxastic faculty as well as the faculty responsible for belief about the phenomenal character of visual experiences. The perceptual and the phenomenological visual faculties share the eye as a component but

---

[24] In the next four paragraphs of the text, I draw heavily on Gould (1991), Boyer (1994; 2002) and Boyer and Barrett (2005).

each draws on it differently, for it enables the formation of belief with different kinds of intentional content. Provided that they draw on them differently (and thus have different constituent abilities), two faculties may conceivably share *all* of their component mechanisms.

The third thing to note is a consequence of how the first two things are related. Put differently, the first point is that our cognitive and perceptual mechanisms constitute a reservoir of abilities which may be exploited by different types of selection pressures differently in the evolution of faculties. Since a faculty may have multiple components and a mechanism may be a component of multiple faculties, doxastic faculties and their component mechanisms may have a many-to-many relationship. The third point is that it is precisely because a faculty's output belief is the result of engaging a *unique* set of abilities that faculties differ in subject matter in the first place. For instance, the difference in content between phenomenological beliefs and perceptual beliefs can only be due to the exercise of distinct, though surely to some extent overlapping, sets of abilities.

Now we are ready to see why a truth-orthogonally evolved faculty must exhaust our resources for belief about its domain. Since under truth-orthogonal selection belief is pressed into service regardless of its correlation with the truth, the faculty in question emerges because its constituent abilities just luckily happened to be available right when their joint appropriation was adaptive. That is, in the truth-orthogonal evolution of a faculty our ancestors were simply lucky to possess the set of abilities with which such a faculty is identical and to find themselves in a situation which triggered their joint exercise (and thus led to its characteristic output belief) and their contemporaries were unlucky in that they never met these conditions (and thus never benefited from this sort of belief informing their behavior). Truth-orthogonal hypotheses explain, not only the origin of a certain faculty, but also the origin of belief with a certain kind of subject matter. So, any belief with this subject matter *must* involve the activity of the truth-orthogonally evolved faculty, since otherwise the "alternatively" formed belief wouldn't be about the same domain. In the truth-orthogonal case, self-correction is impossible because, contrary to what is commonly supposed, we have no other way of forming belief about the domain.

Furthermore, we can't overcome this limitation culturally, either. Recall the difficulty faced by drag argumentation: since both genetic and cultural factors influence the development of a truth-dependently evolved faculty, while negative drag may *initially* play a role in determining its genetic

correlates, cultural selection may as a result of gene-culture co-evolution counteract it. But, in truth-orthogonal debunking, since the target faculty is *un*reliable, not error-prone, *there is no distortion to counteract*. However the species-wide delusion is culturally elaborated across space or time, the output remains epistemically destitute. Even if our doxastic faculties are extremely culturally malleable, since it doesn't matter which way the cultural winds blow, the challenge from truth-orthogonal explanation remains entirely unabated.

A truth-orthogonal explanation implies that the target faculty comprises our sole resource for belief about its domain. Our epistemic situation with respect to the target facts is an impossible quagmire from which neither knowledge nor justification is attainable because we lack the cognitive reach.

1.3.4 The so-called evolutionary debunking arguments

I have already argued that Street and Joyce's debunking arguments cannot recruit negative drag explanations, but can they somehow be understood to be drift arguments? I'll begin with Street. In "A Darwinian Dilemma for Realist Theories of Value," she argues that value realists, who hold that "there are at least some evaluative facts or truths that hold independently of all our evaluative attitudes" (p. 110), face a dilemma in light of the fact that there is a "striking coincidence" between the evaluative judgments we take to be true and the evaluative judgments evolutionary forces would have plausibly lead us to form: either value realism is true and there is no account of this coincidence or the best account of it available is incompatible with value realism. Street suggests that we should opt for the latter horn and therefore abandon realist theories of value because the former option is unacceptably skeptical, since left unexplained – as it must remain if value realism is true, according to Street – the coincidence is simply "incredible" (p. 125).

Despite frequently conflating negative drag with truth-orthogonal selection, I think it would be fair to say that Street intends to give a drift argument. Street writes,

> The key point to see about [the first horn] is that if one takes it, then the forces of natural selection must be viewed as a purely distorting influence on our evaluative judgements, having pushed us in evaluative directions that have nothing whatsoever to do with the evaluative truth. [...]

> If we take this point and combine it with the [premise] that our evaluative judgements have been tremendously shaped by Darwinian influence, then we are left with the implausible skeptical conclusion that our evaluative judgements are in all likelihood mostly off track, for our system of evaluative judgements is revealed to be utterly saturated and contaminated with illegitimate influence. (2006, pp. 121-122)

Setting aside the apparent allusions to negative drag in her use of "distorting," "off track," and "contaminated," Street's thought seems to be that given value realism there is nothing to anchor the evolution of the evaluative doxastic faculty to the truth and such a connection, she assumes, would be needed to learn the evaluative truth on the assumption of value realism. But from her writings, it is not clear exactly how the lack of this connection casts doubt on our reliability about evaluative matters generally, for we simply do not know enough about how "the forces of natural selection" affected the evolution of evaluative cognition.

As I mentioned above, some evolutionary biologists take the reconstruction of the evolution of human cognition to be beyond scientific reach. While her premise that "our evaluative judgements have been tremendously shaped by Darwinian influence" may be plausible, we do not know it to be true in great enough detail to carry out the truth-orthogonal debunking of morality. Furthermore, even if we did know this, and we could undermine value realism as a result, it does not follow that Street's anti-realist view that "evaluative facts or truths are a function of our evaluative attitudes" would successfully preserve our claim to evaluative knowledge (2006, p. 152). To justify the self-ascription of evaluative knowledge, one must also be able to reliably self-ascribe evaluative beliefs, for to justifiably believe that one knows that *p* one must justifiably believe that one holds the belief that *p*. The question then arises whether we have reason to think that there was an evolutionary advance in which our ancestors acquire the capacity to correctly self-ascribe these beliefs. To preserve the claim to evaluative knowledge in the face of her own Darwinian Dilemma, Street must still explain how and why we have this capacity.

Lastly, even if Street can meet this secondary evolutionary challenge, it is unclear whether evaluative knowledge as she understands it is desirable or even genuinely knowledge. While Street is clear that on her version there is "room for the possibility of evaluative error" (p. 152), that does not make room for the possibility of unreliability, as opposed to mere error-proneness. Evaluative knowledge

would be more easily attained than seems intuitively acceptable. When we ascribe moral knowledge, for example, we tend to think that our evidence of the ascription of this knowledge far outstrips our evidence for ascribing the correlate belief. Street never explains why her conception of evaluative knowledge isn't itself already too much of a concession to the skeptic.

Unlike Street, Richard Joyce (2001; 2006; 2016) does not think that realism about the target domain (which for him is just the moral domain) is necessary to arrive at the skeptical conclusion. Like Street, he thinks a credible case can be made that selectionist considerations imply "moral judgments are the output of a non-truth-tracking process" (forthcoming, p. 2). He takes the centrally significant consideration to be that, according to evolutionary accounts of morality, moral judgments need not be true to perform their adaptive function. Presumably, the thought is that moral facts aren't to figure in the explanation in any capacity, so he would also claim that they need not be false, either. So far there appears to be a drift argument on his mind.

But when Joyce attempts to clarify the mechanism by which these considerations debunk he writes things incompatible with the drift interpretation. In particular, he thinks the conclusion to draw is that all moral judgments lack justification but not permanently so (forthcoming). However, if I'm right and a drift argument tell us that the target faculty has the epistemic pedigree of a Magic 8-Ball, then the truth-orthogonal explanation of moral cognition does establish that moral judgments lack justification permanently. In explaining the hedge, he writes that

> the thesis that seems correct to me is that when the belief is formed as a knee-jerk default, without reflection or proper sensitivity to the available evidence—when, that is, the belief is to be explained largely by reference to the arousal of an evolved non-truth-tracking doxastic faculty—then it lacks epistemic justification. (ibid., 9; emphasis in original)

Joyce here equates belief formed without reflection or proper sensitivity to the available evidence with that formed by an evolved "non-truth-tracking" faculty. But such a faculty, then, cannot be taken to be one that evolved truth-orthogonally, for at least two reasons. First, the epistemic defect in the truth-orthogonal case isn't that it's improperly sensitive to the available evidence, it is that on account of it one takes truth-irrelevant considerations to be evidentially relevant. If the faculty did indeed evolve truth-

orthogonally, whatever evidence there is on the matter (if any) is certainly not "available" to us. Rather, we would simply be highly susceptible to the illusion of its availability. Second, contrary to what he suggests, no amount of serious thought or consideration will suffice to correct this epistemic deficiency because no reflection on the relevant matters would be suitably independent of the truth-orthogonally evolved faculty. Truth-orthogonal selection does not construct faculties with the features he describes.

In truth, the faculty Joyce describes can only be the product of negative drag – just look at what he goes on to claim immediately after the above-quoted passage:

> But one is not necessarily stuck in that position [of lacking justification]. We are creatures with the capacity to bring other psychological faculties to bear on the matter—faculties that can track the truth in a reliable manner—and when these are employed properly, the same belief that was once unjustified may become justified. (ibid.)

Only if the faculty evolved by negative drag will one not be stuck with the unjustified beliefs. This will necessarily be the case in the truth-orthogonal alternative. It is only in cases of negative drag that there will be a background of reliability on which one may rely in the manner Joyce describes, that is, against which one may check for the distorting, error-inducing influence of the evolved "non-truth-tracking" processes. But one cannot draw the domain-wide skeptical conclusion precisely because this kind of self-correction presupposes that the error is suitably local.

While Joyce seems to think the moral doxastic faculty must have been truth-orthogonally selected, he confuses the result of this kind of selection with the product of negative drag, leading him to draw the wrong conclusion. This confusion may be due to one of two mistakes (or perhaps both): first, he might think, alongside many anti-debunkers, that it makes sense to claim that there are alternative means for belief about a domain whose faculty evolved truth-orthogonally and/or, second, he might think that negative drag has domain-wide implications. Whatever the case, I think the failure to distinguish negative drag from truth-orthogonal selection is the ultimate source of the error as it is in so much of the literature. Until the error is corrected, one can't say whether he indeed aims to give a drift argument or not.

1.4 Conclusion

I have distinguished between truth-dependent and truth-orthogonal adaptationist explanation and among the former between positive and negative drag explanation. I have argued that if we learn that a truth-orthogonal or a positive drag or a negative drag explanation is correct, we learn, respectively, that the target doxastic faculty has either never yielded justified belief, usually yielded justified belief, or sometimes yielded unjustified belief. Furthermore, I have argued that the rational accommodation of this sort of evidence would, respectively, require one to either permanently abandon the output beliefs of the target faculty, increase one's credence in their justification, or decrease it, perhaps to the point of defeating their pre-existing justification, if they happen to be justified.

Along the way, in the light of this account, I reviewed recent metaethical debunking arguments and suggested that, due to a variety of factual and conceptual errors, all are either unsound or non-evolutionary. I advanced as the most significant error the conflation of truth-orthogonal selection with negative drag, as it is responsible for much of the mismanagement of the core epistemological issues in the work of the two most influential contributors. In addition to relevant empirical and scientific detail, the debunking debates in ethics have largely neglected the epistemological nitty-gritty and, likewise, to the detriment of parties to the discussion on all sides.

EVOLUTIONARY ARGUMENT	TYPE	EPISTEMIC UPSHOT	PROPONENTS
Positive Drag	Buttressing	Justification Reaffirmer	Evolutionary Boosters
Negative Drag	Debunking	Justification Underminer	Modest Debunkers
Drift	Debunking	Justification Blocker	Ambitious Debunkers

Table 1. Summary of conclusions.

References

Abbey, A. (1982). Sex differences in attributions for friendly behavior: Do males misperceive females' friendliness? *Journal of Personality and Social Psychology*, 830-838.

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences*, 29-34.

Bedke, M. (2014). No coincidence? *Oxford Studies in Metaethics, IX*, 102-125.

Berker, S. (2014). Does evolutionary psychology show that normativity is mind-dependent? In J. D'Arms, & D. Jacobson (Eds.), *Moral psychology and human agency: Essays in the new science of ethics.* Oxford: Oxford University Press.

Boyer, P. (1994). Cognitive constraints on cultural representations: Natural ontologies and religious ideas. In L. A. Hirschfeld, & S. A. Gelman (Eds.), *Mapping the mind* (pp. 391-411). Cambridge: Cambridge University Press.

Boyer, P. (2002). *Religion explained: The evolutionary origins of religious thought.* New York: Basic Books.

Boyer, P., & Barrett, H. C. (2005). Domain specificity and intuitive ontology. In D. M. Buss (Ed.), *Handbook of evolutionary psychology* (pp. 96-118). Hoboken: John Wiley and Sons, Inc.

Brosnan, K. (2011). Do the evolutionary origins of our moral beliefs undermine moral knowledge? *Biology and Philosophy*, 51-64.

Christensen, D. (2010). Higher-order evidence. *Philosophy and Phenomenological Research*, 185-215.

Clarke-Doane, J. (2012). Morality and mathematics: The evolutionary challenge. *Ethics*, 313-340.

Clarke-Doane, J. (2015). Justification and explanation in mathematics and morality. (R. Shafer-Landau, Ed.) *Oxford Studies in Metaethics, X*, 80-103.

Copp, D. (2008). Darwinian skepticism about moral realism. *Philosophical Issues*, 186-206.

Crisp, R. (2006). *Reasons and the good.* Oxford: Oxford University Press.

Darwin, C. (1859/2009). *The origin of species by means of natural selection the preservation of favoured races in the struggle for life* (150th anniversary ed.). New York: Signet Classics.

Dawkins, R. (1982). *The extended phenotype: The long reach of the gene.* Oxford: Oxford University Press.

Deem, M. J. (2016). Dehorning the Darwinian dilemma for normative realism. *Biology and Philosophy*, 727-746.

Dennett, D. C. (1987). *The intentional stance.* Cambridge, MA: MIT Press.

Dretske, F. (1989). The need to know. In M. Clay, & K. Lehrer (Eds.), *Knowledge and skepticism* (pp. 89-100). Boulder, Colorado: Westview Press.

Enoch, D. (2010). The epistemological challenge to metanormative realism: How best to understand it, and how to cope with it. *Philosophical Studies*, 413-438.

Fisher, R. A. (1930). *The genetical theory of natural selection.* Oxford: The Clarendon Press.

FitzPatrick, W. (2008, December 19). *Morality and evolutionary biology*. Retrieved from Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/entries/morality-biology/

FitzPatrick, W. (2014). Debunking evolutionary debunking of ethical realism. *Philosophical Studies.* doi:10.1007/s11098-014-0295-y

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: The MIT Press.

Fraser, B. (2014). Evolutionary debunking arguments and the reliability of moral cognition. *Philosophical Studies*, 457-473.

Godfrey-Smith, P. (2000). On the theoretical role of "genetic coding". *Philosophy of Science*, 26-44.

Godfrey-Smith, P. (2007). Conditions for evolution by natural selection. *Journal of Philosophy*, 489-516.

Godfrey-Smith, P. (2007). Information in biology. In D. Hull, & M. Ruse (Eds.), *The Cambridge companion to the philosophy of biology* (pp. 103-119). Cambridge: Cambridge University Press.

Gould, S. J. (1991). Exaptation: A crucial tool for an evolutionary psychology. *Journal of Social Issues*, 43-65.

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

Greene, J. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The neuroscience of morality* (pp. 35-79). Cambridge, MA: The MIT Press.

Haselton, M. G. (2003). The sexual overperception bias: Evidence of a systematic bias in men from a survey of naturally occurring events. *Journal of Research in Personality*, 34-47.

Hawthorne, J. (2004). *Knowledge and lotteries.* Oxford: Oxford University Press.

Hawthorne, J. (2013). The case for closure. In M. Steup, J. Turri, & E. Sosa (Eds.), *Contemporary debates in epistemology* (2nd ed., pp. 26-42). Wiley Blackwell.

Horowitz, S. (2013). Epistemic akrasia. *Nous*, 718-744.

Huemer, M. (2008). Revisionary intuitionism. *Social Philosophy and Policy*, 368-392.

Johnson, D. (2015). *God is watching you: How the fear of God makes us human.* New York: Oxford University Press.

Joyce, R. (2001). *The myth of morality.* Cambridge: Cambridge University Press.

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

Joyce, R. (2016). Evolution, truth-tracking, and moral skepticism. In R. Joyce, & B. Reichardt (Ed.), *Essays in moral skepticism* (pp. 142-158). Oxford: Oxford University Press.

Kahane, G. (2011). Evolutionary debunking arguments. *Nous*, 103-125.

Kitcher, P. (2011). *The ethical project.* Cambridge, MA: Harvard University Press.

Kitcher, P. (2016). Evolution and ethical life. In D. Livingstone Smith (Ed.), *Biophilosophy.* Cambridge: Cambridge University Press.

Lewontin, R. C. (1998). The evolution of cognition: Questions we will never answer. In *An invitation to cognitive science: Methods, models, and conceptual issues* (Vol. IV, pp. 107-132). Cambridge, MA: The MIT Press.

Liu, J., Harris, A., & Kanwisher, N. (2010). Perception of face parts and face configurations: An fMRI study. *Journal of Cognitive Neuroscience*, 203-211.

Mark, J. T., Marion, B. B., & Hoffman, D. D. (2010). Natural selection and veridical perceptions. *Journal of Theoretical Biology*, 504-515.

McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, 493-510.

Millikan, R. G. (1984). Naturalist reflections on knowledge. *Pacific Philosophical Quarterly*, 315-334.

Mogensen, A. (2015). Do evolutionary debunking arguments rest on a mistake about evolutionary explanations? *Philosophical Studies*, 1-19.

Nagel, T. (1979). Ethics without biology. In *Moral questions* (pp. 142-146). Cambridge: Cambridge University Press.

Nagel, T. (2012). *Mind and cosmos: Why the materialist Neo-Darwinian conception of nature is almost certainly false.* Oxford: Oxford University Press.

Nozick, R. (1981). *Philosophical explanations.* Cambridge, MA: Harvard University Press.

Nozick, R. (2001). *Invariances: The structure of the objective world.* Cambridge, MA: Havard University Press.

Parfit, D. (1984). *Reasons and persons.* Oxford: Oxford University Press.

Parfit, D. (2011). *On what matters.* Oxford: Oxford University Press.

Pritchard, D. (2007). *Epistemic luck.* Oxford: Oxford University Press.

Quine, W. (1975). The nature of natural knowledge. In S. Guttenplan (Ed.), *Mind and language* (pp. 67-81). Oxford: Clarendon Press.

Richardson, R. C. (2007). *Evolutionary psychology as maladapted psychology.* Cambridge, MA: MIT Press.

Ruse, M. (1986). *Taking Darwin seriously: A naturalistic approach to philosophy.* Oxford: Basil Blackwell.

Setiya, K. (2012). *Knowing right from wrong.* Oxford: Oxford University Press.

Shafer-Landau, R. (2003). *Moral realism: A defence.* Oxford: Oxford University Press.

Singer, P. (2005). Ethics and intuitions. *The Journal of Ethics*, 331-352.

Skarsaune, K. O. (2011). Darwin and moral realism: Survival of the iffiest. *Philosophical Studies*, 229-243.

Sterelny, K. (2003). *Thought in a hostile world: The evolution of human cognition.* Malden, MA: Blackwell Publishing.

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

Street, S. (2008). Reply to Copp: Naturalism, normativity, and the varieties of realism worth worrying about. *Philosophical Issues*, 207-228.

Stroud, B. (2002). Evolution and the necessities of thought. In *Meaning, understanding, and practice: Philosophical essays* (pp. 52-66). Oxford: Oxford University Press.

Trivers, R. L. (1976). Foreword. In R. Dawkins, *The selfish gene.* Oxford: Oxford University Press.

Vavova, E. D. (2014). Debunking evolutionary debunking. *Oxford Studies in Metaethics, IX*, 76-101.

White, R. (2010). "You just believe that because ...". *Philosophical Perspectives*, 573-615.

Wielenberg, E. (2010). On the evolutionary debunking of morality. *Ethics*, 441-464.

CHAPTER 2

The Evolutionary Challenge and the Evolutionary Debunking of Morality

We ordinarily think we know right from wrong, good from evil, and quite often how we morally ought, or ought not, to act. Recently, some philosophers have appealed to our Darwinian origins in an attempt to argue that we do not have any moral knowledge.[25] I will argue that our origins do pose an epistemological challenge, but that it derives instead from what we don't know about our evolutionary history. I argue that we don't know enough about our Darwinian past to know that we ever evolved the *capacity* to acquire moral knowledge. In contrast to recent metaethical debunking arguments, I conclude that it is what we *don't* know about our evolutionary past, not what we do know, that raises the truly formidable challenge.

The evolutionary challenge from ignorance differs from all others on several fronts. First, unlike all others known to me, the challenge here is not to *explain* our reliability about moral matters, but to justify our claim to moral knowledge *in the absence of any empirical evidence for our reliability*. I argue that

---

[25] Joyce (2001; 2006; 2016) is the most well-known proponent. Some, e.g., Fraser (2014), have argued on evolutionary grounds that it is not reasonable to expect moral cognition to be reliable. Others, e.g., Street (2006), argue that skepticism follows on the assumption of certain forms of metaethical realism.

since we don't have the evidence to think that we have reliable moral cognition we don't know moral knowledge exists. Second, unlike most other challenges, this one doesn't assume moral cognition evolved adaptively. I argue that the leading selectionist hypothesis is incompatible with the existence of moral knowledge and that it is its status as *uneliminated*, rather than true or substantiated, that presents a skeptical challenge. Third, unlike Street's (2006) and Ruse's (1986) arguments, the present argument does not merely target moral realism. For it relies on domain-general epistemic principles and the assumption of Darwinian naturalism: the conjunction of *Common Descent*, the thesis that every living thing on Earth is related by descent with modification, and *Selectionism*, the view that natural selection has been the main, but not exclusive, means of modification.[26] Lastly, I try to show that we lack the justification for thinking there is human moral knowledge, not to establish that we are in fact morally ignorant. I will explain how despite the logically weaker conclusion this challenge is dialectically on par with any attempt to directly show that there is no moral knowledge.

The plan for this chapter is as follows. First I describe the conditions under which moral cognition would have evolved entirely without regard to the moral truth (§2.1). Then I argue that such a *truth-orthogonal* descent, as I will call it, is incompatible with the existence of human moral knowledge (§2.2). I then go on to argue that since we lack the evidence to rule out its truth-orthogonal descent we lack the evidence to know humans evolved the capacity to acquire moral knowledge (§2.3). I conclude that, on the assumption of Darwinian naturalism and the closure of knowledge under known entailment, we don't know whether we can acquire moral knowledge. I suggest to end that in the absence of the knowledge that there is moral knowledge moral skepticism may be a reasonable solution to the problem of moral knowledge.

2.1 A Truth-Orthogonal Explanation of Moral Cognition

In this section, I develop a scientifically informed account of the evolution of morality on which moral cognition evolved entirely without regard to the moral truth, both realistically construed and not.

---

[26] The *locus classicus* of Darwinian naturalism is, of course, *On the Origin of Species*. For philosophically, scientifically, and historically informed discussion of Darwinian naturalism, see Hodge & Radick (2009).

As we will see, it doesn't immediately follow that such a truth-orthogonal descent is incompatible with the existence of moral knowledge, but I go on to describe in §2 the conditions under which it would be so. Put in the abstract, the evolutionary challenge from ignorance just is the challenge to defend, or argue for, the view that either moral cognition does not have a truth-orthogonal descent or that, if it does, the conditions that would make such a descent incompatible with moral knowledge do not obtain. I fill in the details needed to make it compelling in what follows.

2.1.1 The moral prosociality hypothesis

Most moral evolution theorists propose accounts on which moral cognition evolved to play a role in the evolution of human cooperation. They, e.g., Tomasello (2016), argue that the capacity for moral cognition emerged as a by-product of psychological capacities which were recruited over subsequent generations to secure large-scale cooperation among genetically unrelated individuals. Since cooperation involves the coordination of behavior towards a common goal, and moral beliefs are precisely about *how we ought to behave*, the content of moral beliefs aids in cooperation by prescribing courses of action that in that environment would enable and promote helping and sharing behavior and deter cheating, defection, and other uncooperative behaviors. This is the idea at the core of (almost) every moral evolution theory out there. I call this *the moral prosociality hypothesis*.

Different evolutionary theorists of course appeal to different mechanisms. Following Kitcher (2016), we can distinguish the *Selectionist* and *Genealogical* traditions in approaching the relation between evolution and morality. The first, familiar from the metaethical literature, takes Darwin's notion of natural selection to be the crucial explanatory connection. One prominent proponent advocates the thesis that "human morality is a distinct adaptation wrought by biological natural selection" (Joyce, 2008). The second approach, Kitcher's own in *The Ethical Project*, aims to show, by drawing geneological connections, how human moral psychology might have emerged from non-human capacities, leaving the action of

natural selection in the background. The main aim is to describe a set of transitions that trace the development of moral cognition from the likely traits of our non-moralizing ancestors.[27]

On both sides of the divide, to earn their evolutionary keep, our ancestors' moral (or premoral)[28] beliefs just had to be the most prosocial of the variants. The traditions differ in *how* these beliefs earn their keep across the generations. The Selectionist claims that, relative to that of their contemporaries, our ancestors' moral beliefs caused, or causally correlated with, greater reproductive success, eventually out-reproducing non-kin peers. This selective advantage is thus assumed to have a genetically transmittable basis.[29] By contrast, the Genealogical approach takes the mode of transmission to be primarily cultural. Our ancestors' moral belief system gained enough adherents, or converts, to crowd out alternatives. The spread of moral beliefs takes place both within and across generations. Whatever the mode of transmission, however, all moral evolution theorists agree that morality as we know it owes its existence to the fact that our ancestors' moral beliefs happened to have the content that, in their environment, best secured large-scale cooperation among the variants.

> More precisely, the moral prosociality hypothesis states that
>
> Given a set of individuals that exhibited variation in (pre-)moral opinion, those with the most prosocial of the moral beliefs had more babies or more students than those with less prosocial (pre-)moral beliefs.[30]

A belief is *prosocial* iff it enables or promotes helping and sharing behavior and/or deters uncooperative behavior like cheating and defection. Our beliefs about traffic laws, their content, and their violation's

---

[27] This is Darwin's own approach in *The Descent of Man*. For an illuminating discussion of Darwin's genealogical account, as well as an excellent critique of the Selectionist approach, see Kitcher (2016).

[28] By "premoral beliefs" I mean the proto- or pseudo-moral psychological precursors to full-blown moral beliefs. For brevity, I omit the qualification in what follows, but it is always intend it where relevant.

[29] Kahane (2011) notes this assumption but, implicitly working with the Selectionist assumptions, he incorrectly takes this to be an assumption all evolutionary debunking arguments must make (see p.112). There's no reason to think that the Genealogical approach has any less skeptical potential than the Selectionist one.

[30] I take the description of fitness as measured by "having babies" and "having students," from Sober (2006), where he attributes the latter to cultural evolution theorist Peter Richerson. For simplicity, as in Chapter 1, I assume the type of selection involved is directional.

cost are prosocial to the degree that these aid in the coordination of transportation. The hypothesis states that however the transmision of moral belief is best understood the main determinant of the content of morality as we know it has been individual differences in the prosociality of past moral cognition.

The moral prosociality hypothesis makes two noteworthy assumptions. First, it assumes that moral beliefs helped causally generate behavior. *Pace* Street (2006; 2008), this assumption is orthogonal to the internalism-externalism debate about moral motivation. It is beside the point whether our moral beliefs have an intrinsic connection to motivation or action, that is, whether they cause behavior on their own or alongside desire.[31] They are simply hypothesized to play a causal role in the evolution of cooperation among non-kin. Second is the assumption that the phenomenon of moral cognition is unified enough to warrant special explanation. There is something in virtue of which all moral beliefs are moral rather than, say, mathematical, physical, biological (etc.), and this marks it off as calling for separate explanation. This is usually taken to be its distinctive subject matter.

Intuively, the moral prosociality hypothesis does appear to severely threaten moral knowledge. Unlike your beliefs about traffic laws, in moral belief accuracy and prosociality don't seem to reliably co-vary. Indeed, morality resembles religion in this respect. If you believe God will punish you if you defect or cheat, you're likely to avoid acting uncooperative regardless of whether God exists. The same seems true of many moral beliefs, such as the belief that parents have a special moral obligation to protect, and care for, their children or that one should avoid harming others and help those in need. I think that we perceive a skeptical threat to morality in the fact of evolution because many of us are already quite skeptical of religion on historical grounds.

In coming sections, I explore the analogy with religion to understand how our evolution might create trouble for our pretensions to moral knowledge. But first, I try to understand how, in general, the evolution of a kind of cognition may be wholly indiferent to its purported objects of knowledge.

---

[31] Berker (2014) makes a similar point in his fn. 26.

2.1.2 Truth-dependent vs. truth-orthogonal selection

Human cognition with respect to a certain subject matter could in principle have been selected truth-dependently or truth-orthogonally. Roughly, *truth-dependent selection* occurs when individual differences in reliability are causally correlated with different degrees of fitness and *truth-orthogonal selection* occurs when it is individual differences in cognition *other than reliability* that are causally correlated with different degrees of fitness. For differences across members of a population to be *causally correlated with* different degrees of fitness just is for these differences to causally contribute to, or have a causally relevant factor in common with, lifetime reproductive and/or pedagogical output, where these are measured by absolute number of, respectively, babies and students.

This distinction is neutral on whether the mode of transmission is primarily cultural or biological, since fitness is measured by either number of babies or mostly[32] number of students. In consequence, as drawn, the distinction cuts across both natural and cultural selection. If cognition about a certain subject matter is wholly the result of truth-orthogonal selection, either as an adaptation or its by-product,[33] it has a *truth-orthogonal descent*; if it is at least partially the result of truth-dependent selection, either as an adaptation or its by-product, it has a *truth-dependent descent*.

One may try to explain the existence of a certain kind of cognition by positing either type of descent. We can distinguish between truth-dependent and truth-orthogonal hypotheses accordingly. On truth-dependent hypotheses, our ancestors reproduced and/or taught more prolifically than their same-species contemporaries because they were, across some segment of our evolutionary history, more (or less) reliable about the subject matter in question than their contemporaries. On truth-orthogonal hypotheses, by contrast, our ancestors outreproduced or outproselytized their contemporaries because, irrespective of reliability, they formed beliefs that, in that environment, were more adaptive than that of

---

[32] Recall that on the Genealogical approach natural selection does have background explanatory role.

[33] Adaptations and their by-products are biological or cultural depending on whether they are the result of natural or cultural selection.

their contemporaries. These two types form a dichotomy of the class of selectionist (as opposed to non-selectionist)[34] cultural and biological evolutionary hypotheses.

It is easiest to appreciate the distinction with examples. Many cognitive anthropologists, e.g., Johnson (2015), have argued that the capacity for supernatural or religious belief first emerged as a by-product of distinct psychological capacities which were recruited over subsequent generations to help secure cooperation among non-kin. They hypothesize that belief in the proximity of punitive supernatural moralizers increases prosocial behavior by creating the impression of being watched when alone and by inducing the fear of supernatural punishment for behaving uncooperatively. Since on this hypothesis natural selection proceeded in complete indifference to the supernatural facts, and individual differences in reliability about supernatural matters are not hypothesized to be either selected for or selected against, *the supernatural punishment hypothesis*, as it is called, posits the truth-orthogonal *biological* descent of supernatural cognition. Once we add to this hypothesis that there have only been false prophets, then, supernatural belief would have a completely truth-orthogonal descent.[35]

Truth-orthogonal hypotheses, however, are very different from truth-dependent hypotheses that involve selection for *error-prone* cognition.[36] For insnce, according to certain evolutionary psychologists, e.g., Haselton (2003), our ancestors reproduced more prolifically than their contemporaries due in part to the so-called *male sexual overperception bias*, the male tendency to overattribute sexual interest to women based on friendliness. They hypothesize that the bias made our forefathers less likely to overlook sexually receptive females.[37] (I am not endorsing this hypothesis, just reporting it.) Such *adaptive bias* hypotheses

---

[34] Non-selectionist hypotheses include those that appeal to mechanisms of genetic drift or mutation. While unlikely these mechanisms could conceivably explain the evolution of moral cognition. The argument of this chapter, as will become clear later in the text, doesn't hinge on assuming a selectionist hypothesis is the correct explanation of the evolution of moral cognition, but just on the claim that such a hypothesis might be correct.

[35] What if most of our supernatural beliefs are true and a third factor explains both the supernatural facts and our supernatural beliefs' accuracy? Then either the explanation of the correlation appeals to the action of a further non-natural agent, flouting Darwinian naturalism by substituting natural selection with artificial selection, or the descent of supernatural cognition is truth-dependent, since individual differences in reliability would be causally correlated with relative fitness.

[36] It is standard fare to confuse the two. As Kahane (2011) writes, "it is now common to think of [evolution] as a *distorting* influence on our evaluative beliefs [...]. And implicit or explicit debunking arguments that rely on this assumption are fairly widely used in contemporary normative ethics" (p. 109; emphasis in original).

[37] I write "females" rather than "women" because the bias may predate the advent of modern humans.

always posit a truth-dependent descent. For, prior to the initial development of the bias, selection on our lineage *must have been* truth-dependent because the adaptiveness of the bias is contingent on a background of true belief (McKay & Dennett, 2009). To see this, just notice how if our forefathers had instead attributed sexual interest to rocks or plant life or other animal species or same-species males or same-species menopausal females, the bias would obviously not have been adaptive. Our forefathers must have been reliable enough to correctly identify the targets of *adaptive* misattribution, same-species females of child-bearing age. Contrast this reliance on the facts with facts' complete irrelevance in truth-orthogonal explanation. On the supernatural punishment hypothesis, the supernatural agents can be believed to be any old thing provided it enables cooperation among non-kin.

I hope the distinction is by now clear enough. On truth-dependent hypotheses, individual differences in reliability causally correlate with different degrees of fitness and on truth-orthogonal hypotheses it is differences in cognition other than degree of reliability, e.g., degree of prosociality, that causally correlate with these differences in fitness. In the next section, I argue that the moral prosociality hypothesis is truth-orthogonal because moral belief is not belief about prosocial matters.

2.1.3 A condition on truth-orthogonality

The moral prosociality hypothesis states that moral beliefs waxed and waned in popularity according to relative prosociality. Some moral beliefs may have been too anti-social or maladaptive to have been held by any of us for very long. Perhaps some individuals were so fervently environmentalist as to be driven to suicide by their uncompromising conservationist stance. Other moral beliefs may have been held for thousands of years, prosocial enough for the untold, prehistorical savagery of the distant past, but far too inimical to cooperation at the cradle of human civilization. Perhaps a remnant of such a past is the barbaric doctrine of *lex talionis*, a principle developed in early Babylonian law on which, e.g., proper punishment for killing a peer's daughter is to have your own daughter killed. Yet other moral beliefs may be so essential to the functioning of human society that no circumstance could enduringly upset them. Perhaps our felt conviction that members of our in-group are beings with moral significance, deserving of fair treatment and moral concern, can never be upended. Among the cultural oceans of humanity the sea of morality has seen its gait set by the powerful currents of prosociality.

The moral prosociality hypothesis posits a truth-orthogonal descent on the condition that the truth-conditions of moral beliefs are not identifiable with, reducible to, or even reliably coincident or correlated with facts about the circumstances that causally determined present-day moral demographics – in slogan form, that *moral beliefs are not about prosocial matters*. Consider that on this condition that this hypothesis does not require or imply that any of our ancestors *ever* held any true moral beliefs. For all it says, the true moral beliefs never became popular for the simple reason that no one has ever held them. So, even if we suppose that moral belief may be true, and that the true moral beliefs are the most prosocial, the moral beliefs we have today may nevertheless all be false, for neither truth nor probability follows from, or is otherwise guaranteed by, currently unsurpassed prosociality. Moreover, even if true moral beliefs were in fact held by our ancestors, and these are also the most prosocial of the variants, on the moral prosociality hypothesis they were selected entirely due to their prosociality, accuracy be damned. Had our ancestors' contemporaries held *false* moral beliefs that were *more* prosocial, all else the same, *their* descendants would be alive today, not us.[38]

If moral beliefs are not about prosocial matters, the moral prosociality hypothesis posits the truth-orthogonal descent of moral cognition because the truth-conditions of moral beliefs would not be identifiable with, reducible to, or reliably correlated with facts about the circumstances that causally generated human morality as we know it. In what follows, I will argue that because we do not know the moral prosociality hypothesis is false and we do not know moral beliefs are about prosocial matters that we do not know there is human moral knowledge. In the next section, I identify the condition on which the moral prosociality hypothesis is relevant to moral epistemology.

2.1.4 A Psycho-Moral Nexus?

The moral prosociality hypothesis, some seem to think, is only relevant to the question of moral knowledge if we assume our moral beliefs and the moral facts are constitutively independent. As I will

---

[38] More precisely, we would not be around today if our ancestors failed to convert to the moral belief system of their contemporaries.

show in the next section, this is not quite right but it does show that evolutionary etiology may, on some metaethical accounts, be entirely immaterial to the explanation of our reliability.

On certain meta-ethical views, e.g., Street (2008), moral cognition is constitutively bound up with its subject matter. Our attitudes are said to figure in an account of what it is for an action to be right or wrong, good or evil, permissible or impermissible, etc. Focusing on the rightness of actions for simplicity, one such account would be

> SIMPLE CONSTRUCTIVISM: The fact that X is a right-making reason to Y for agent A is constituted by the fact that A judges X to be a right-making reason to Y.

I call "constructivist" views on which the truth of moral propositions is explained in terms of our attitudes to them. This view simply lets the facts about rightness hang on what we in fact judge. In consequence, the descent of moral cognition has no bearing on its reliability, for on either a truth-orthogonal or truth-dependent descent we believe right the actions we judge to be right. On this view, when we come to believe what we judge, we form moral beliefs reliably whatever their evolutionary etiology.

Another view that posits a psycho-moral nexus appeals to the nature of concept possession. To have a certain moral concept, one's application of the concept must be sufficiently reliable. Following Setiya's (2012) formulation, and again focusing on rightness for simplicity, it is

> SIMPLE EXTERNALISM: Part of what it is to have the concept of moral rightness is to be such that one's method for identifying actions as right is sufficiently reliable.

Setiya calls "externalist" views on which moral concept-possession is explained in terms of the reliability of one's method for moral belief formation. Human moral cognition involves the application of moral concepts iff we are sufficiently reliable about moral matters. Here the constitutive explanation goes in the opposite direction: from moral fact to moral belief. Without moral reality, there can be no moral belief, for there would be no moral concept. As before, the truth-orthogonal descent of moral cognition does nothing to detract from its reliability, since there would be a metaphysical guarantee of our reliability on either descent. If there is a psycho-moral nexus of either an externalist or a constructivist

kind, moral evolution could have proceeded entirely without regard to moral reality yet, nevertheless, yield the capacity to acquire moral knowledge.

In the coming sections, I argue that we do not know whether we have moral knowledge because for all we know (i) the moral prosociality hypothesis is true, (ii) moral belief is *not* about prosocial matters and (iii) there is no psycho-moral nexus.

2.2 The Moral Ignorance Hypothesis

In this section I argue that if the moral prosociality hypothesis is correct, and moral beliefs are not about prosocial matters, *and* there is no psycho-moral nexus, none of us has ever had, or could ever come to acquire, any moral knowledge. But first, a note on debunking.

2.2.1 Two Types of Debunkers

Call the negative epistemic effect of learning about certain causes of one's beliefs *debunking*. Following White (2010), we can distinguish undermining debunkers and blocking debunkers. An *undermining debunker* presents us with etiological information that *defeats* the epistemic standing of the target belief. You're told that the calculator you just used to arrive at a tip amount malfunctions every now and then – whether it misled you just then this information defeats your justification for believing the tip amount is correct. Even if the calculator did in fact function properly, in light of this news, it is natural to think that it would be unjustified for you to hold on to the belief. Your informant here is an undermining debunker.

A *blocking debunker* argues that the causes of the target belief *block* it from having attained a certain positive epistemic standing. You are told that your personal fortune-telling device, your Magic 8-Ball, is a mere novelty toy and that therefore, your informant explains, your 8-Ball-based beliefs are, if true, luckily true. You learn that facts about the circumstances of their formation blocked the beliefs from ever amounting to knowledge. Here your informant is a blocking debunker. The first kind of debunker defeats the positive epistemic standing of the target belief by revealing information about the circumstances of its formation and the second kind argues that facts about the causal predicament prevent the target belief from ever attaining a certain positive epistemic standing.

In the next section, I argue that in a world where we know that there is no psycho-moral nexus, and we know, moreover, that moral beliefs are not about prosocial matters, the evolutionary debunker that appeals to a truth-orthogonal descent would be a blocking debunker. That is, assuming there is no psycho-moral nexus and that moral beliefs are not about prosocial matters, if we were to learn that the moral prosociality hypothesis is true, this discovery would be conclusive evidence for the disturbing fact that moral belief does not amount, has never amounted, and will never amount, to knowledge.

2.2.2 The truth-orthogonal debunking of morality

The logic behind truth-orthogonal debunking of morality is straightforward. If moral cognition has a truth-orthogonal descent, and there is no psycho-moral nexus, there would be no nonaccidental connection between our moral beliefs and their truth, since their selection would have entirely truth-orthogonal and there would be no constitutive connection between two. Moreover, since luckily true belief does not amount to knowledge, moral belief could not amount to knowledge even if true. In consequence, if the moral prosociality hypothesis is true, and moral belief is not about prosocial matters, and there is no psycho-moral nexus, then moral belief has never been knowledge and never will be.

If this sounds unconvincing in the abstract, perhaps an analogy will help. Suppose we learn that supernatural cognition has a truth-orthogonal descent. In line with the supernatural punishment hypothesis from the previous section, our god-fearing ancestors reproduced more prolifically than their nonbelieving contemporaries because belief in punitive supernatural moralizers promoted cooperation in their environment. We owe our existence to the fact that cognition of a specific, fear-inducing, paranoia-inviting kind happened to be adaptive in the environment our ancestors found themselves.

Now, suppose supernatural agents do in fact roam the cosmos, and that Earth catches the eye of some of the punitive moralizers in the bunch. They visit around the time of our split with chimpanzees and bonobos (6-10 million years ago), and stay to observe the course of hominid evolution without ever interferring in any way. Irony of ironies, we develop the capacity for supernatural belief, and even come to form belief in punitive supernatural moralizers, yet never do we develop the capacity to detect any trace of their presence perceptually. To this day the spectating moralizers remain on Earth in the hopes of one day defying hominid perception no more.

With these suppositions locked and loaded, ask yourself: do we presently have the capacity to *know* that there are supernatural moralizers in our midst? The answer seems to be firmly in the negative. Since we lack the means to detect *any* trace of our spectators, we lack the ability to base supernatural belief in the evidentially relevant facts and therefore – as follows from familiar post-Gettier epistemology – we lack the capacity to know the supernatural facts, for the truth of our supernatural beliefs would be accidental, lucky. (Cf. Gettier 1963.)[39] The truth-orthogonal descent of supernatural cognition blocks supernatural belief from ever amounting to knowledge because its conformity to the truth, however precise, is inevitably a matter of accident. Given a truth-orthogonal descent, our supernatural belief, whatever its accuracy, cannot amount to knowledge.

Critically, we are epistemically closed off from our spectators not just because human perception never evolved to register their presence, but more significantly because supernatural cognition evolved truth-orthogonally. Even if we suppose our spectators perceptible we must be able to recognize that what we perceive *indicates their presence*. If our spectators are *in fact*, and *always have been*, what we call "clouds", then while we have plenty of evidence of their presence we can't use this evidence to turn supernatural belief into knowledge because we can't even comprehend *how* what we have been calling clouds are really supernatural moralizers. Some breakthrough would be needed in how we think about clouds to see how these collections of water droplets are even agents, let alone *supernatural* agents, let alone *moralizing* supernatural agents. Why would these things have *any* code of conduct, let alone a moral code?

It should be clear that, since on the new supposition that the spectators are *really* – not in disguise but au naturel – what we have been calling clouds, no manner of mere *perceptual* augmentation will help us see this. Rather, we need to first learn to understand, for instance, how clouds can have minds at all. Since we cannot comprehend how our spectators may be clouds, even if we do happen to infer their presence from our perception of clouds, it is sheer luck that these happen to be the spectators because *for all we can understand* we might have just as easily inferred their presence from the sight of mountains or the whirl of the wind. Because supernatural cognition, we are assuming, evolved to promote cooperation, not supernatural understanding, we cannot turn supernatural belief into knowledge

---

[39] For a recent discussion of epistemic luck, see Prichard (2007).

because, even if there are supernatural facts, we didn't evolve the capacity to discern the natural facts of supernatural relevance.

These lessons all carry over to the moral case. If moral cognition has a truth-orthogonal descent, then our moral beliefs are, if ever true, accidentally true, since we lack the understanding to do anything more than guess at the nonmoral facts of moral relevance. If moral cognition evolved truth-orthogonally, we did not evolve the capacity to see why any action is right or wrong, good or evil, morally permissible or obligatory (etc.). If we ever get any moral question right, it is as a matter of accident that we took the right nonmoral facts to be morally relevant in just the same way that it is an accident that we truly believe that the clouds to be punitive moralizers. In both cases, it is a fluke that we arrive at the truth of the matter. So, given their truth-orthogonal descent, even if most of our actual moral beliefs happen to be true they are not therefore any closer to amounting to knowledge.

I should point out, though, that the lessons generalize to the moral case only if there is no psycho-moral nexus. In the supernatural case, we intuit the absence of a psycho-*supernatural* nexus, and it's clear enough why. The supernatural spectators were present before we ever develop supernatural cognition, and by hypothesis they had no effect on the course of its evolution. Given a truth-orthogonal descent and the absence of any such connection, supernatural belief, if true, would be accidentally true. Given that there is no psycho-moral nexus, that moral belief is not about prosocial matters, and that the moral prosociality hypothesis is correct, we can only conclude that no one – not us, not our ancestors – has ever had any moral knowledge.

2.2.3 Contingency and safety from error

Nevertheless, one might object that the situations are not parallel, even on the assumption that there is no psycho-moral nexus. Part of the explanation for why many of our supernatural beliefs are true is that specifically the punitive supernatural spectators just happened to be attracted to Earth at the dawn of humanity. But, it might be argued, since moral truths are metaphysically necessary, if our moral beliefs are accurate, we could not have evolved to easily be wrong about moral matters, since it is unlikely that the environment would change enough from one generation to the next to render moral beliefs with these contents any less prosocial and so any less adaptive. Since safety from error is usually sufficient for

knowledge, the objection runs, moral belief could still routinely be knowledge, for unlike supernatural belief it may often be safe from error.

While the suggestion that if moral beliefs are in fact accurate moral cognition would be safe from error may be correct, it is a mistake to think that the troublesome contingency depends in any way on the modal status of the target truths. The contingency of the spectators' arrival or their possible departure has nothing to do with why, given a truth-orthogonal descent, supernatural truths cannot be known by us. We can't come to know anything about the supernatural agents because *we did not inherit the capacity to understand supernatural matters well enough to arrive at supernatural knowledge*. It is *we* who pose the epistemological obstacle, not the facts we aim to know, because however the world chips in the target facts, contingently or necessarily, the human mind is constitutively incapable of doing what is needed to come to know them. To see this, suppose further that our supernatural spectators are metaphysically necessary omnipresent beings. Even if this were the case, it would still be an accident *on our end* that we form supernatural beliefs that are true. The contingency of their existence is epistemically irrelevant.

Rather, it is the contingency with which supernatural belief is pressed into adaptive service that seals its epistemic fate. To be successful, a cooperative enterprise must be sensitive to the ecological context that gives rise to the need for cooperation, and there must be cooperators with the capacity to fill their respective mutually supportive roles. The remarkable chanciness of the emergence of belief in *specifically* punitive supernatural moralizers, together with the awesome luckiness of its suitability for the promotion of cooperation *despite the irrelevance of its truth-value*, makes the truth of supernatural belief an accident for the ages. Not only does supernatural belief thereby earn its place in posterity, but for as long as it has that place due to its prosociality its truth is a happy accident, since it plays no role in accounting for its adaptiveness. Their truth-value plays absolutely no role in explaining their popularity.

Importantly, supernatural belief is accidentally true only if the god-fearing *naturally* outreproduce the nonbelieving. If our spectators had, for example, bred us to believe that supernatural agents exist, it is of course no accident that this belief is true, since supernatural agents would have orchestrated its development. The adaptiveness of our ancestors' supernatural beliefs in their environment must be a consequence of the demands placed on them by their natural habitat. A truth-orthogonal descent blocks

belief from amounting to knowledge in part due to its naturalness, since it is precisely the aimlessness of natural selection that makes its truth an accident. This, together with the absence of a psycho-supernatural nexus, entails our utter supernatural ignorance.

This lesson again generalizes to the moral case, for the hypothesized action of natural selection would be just as aimless. Even if most of our moral beliefs are true, and these truths are metaphysically necessary, the fact that we have moral beliefs with these contents obtains because of a mighty heaping of luck, since we would not make enough sense of moral matters to nonaccidentally arrive at the correct answers to moral questions. If moral cognition has a truth-orthogonal descent and there is no psycho-moral nexus, then moral belief is, has always, and will always be, if true, accidentally true, preventing it from amounting to knowledge.

2.3 The Evolutionary Challenge from Ignorance

Call the conjunction of the moral prosociality hypothesis, the view that moral belief is not about prosocial matters, and the claim that there is no psycho-moral nexus, *the moral ignorance hypothesis*. In this section, I argue that because we don't know enough to rule the moral ignorance hypothesis out we don't know if we evolved the capacity to know right from wrong, good from evil, or how we morally ought, or ought not, to act.

2.3.1 The evolutionary challenge and the moral ignorance hypothesis

As the title of the chapter indicates, I distinguish between the evolutionary debunking of morality and the evolutionary challenge. To debunk morality, not only must we must know that there is no such thing as a psycho-moral nexus, we must know moral cognition has a truth-orthogonal descent, but we are nowhere close to knowing any such thing. The evolutionary debunking of morality takes evidence we do not have and cannot expect to ever acquire. Evolutionary debunking arguments do not have the empirical legs to stand on. For this reason, I take the truly formidable challenge from evolutionary considerations to be very different.

As I conceive of it, the evolutionary challenge to our claim to moral knowledge is the challenge to rule out the moral ignorance hypothesis – not that posed by our knowledge of its truth, for we have no such knowledge – and it goes like this:

(P1) We are not in a position to know that the moral ignorance hypothesis is false.

(P2) If we know there is human moral knowledge, we are in a position to know the moral ignorance hypothesis is false.

(C) We do not know there is human moral knowledge.

As explained in §1, if the moral prosociality hypothesis is correct, and moral beliefs are not about prosocial matters, moral cognition descended truth-orthogonally. If moral cognition descended truth-orthogonally, and there is no psycho-moral nexus, then as explained in §2 we would not have moral knowledge. So, if we know we have moral knowledge, we should be able to deduce and thereby come to know that the moral ignorance hypothesis is false. But since we are not in a position to know this, contrary to what we were brought up to believe, we do not know there is human moral knowledge.

We are not in a position to know the moral ignorance hypothesis is false because we aren't in a position to know any of its conjuncts is false. First of all, we clearly lack the empirical evidence to know that the moral prosociality hypothesis is false. Indeed, the available evidence favors its truth and, even if you were bold enough to argue for its falsity, you would run against a few incontrovertible obstacles. Since its denial would be to make a claim about the history of human cognition on Earth, you would need to gather empirical evidence in its favor. But the cognitive traits of past generations leave little or no trace in the fossil record (Richardson, 2007). Indeed, the preservation of such readily decomposing matter as nervous tissue is extraordinarily rare, and to date no hominid specimen has been recovered. As the discovers of the first known dinosaur brain fossil write,

> The soft tissues of vertebrates [...] and terrestrial organisms in particular are [...] rarely preserved. Brain tissues are among the least commonly preserved soft tissues in the fossil

> record because fossilized brains themselves are extremely rare and, more importantly, because most brain tissues are highly labile [i.e., prone to chemical breakdown]. (Brasier, et al., 2016)

The little we know from the fossil record about the evolution of hominid brains comes in the form of fossilized skull fragments and the occasional complete skull, and many early hominid species are just represented by one or a few fossils (Smithsonian Institute, 2016). Fossil and other empirical evidence may even be so inadequate as to proclaim the evolutionary explanation of cognition to be forever beyond empirical reach (Lewontin, 1998). We don't have the evidence to reconstruct the evolution of moral cognition; we are quite clearly not in a position to know that the moral prosociality hypothesis is false.

Second, it is highly implausible that moral beliefs are, as a whole, generically so, *about* prosocial matters, that is, about matters of human cooperation or interpersonal logistics. That an action is right or wrong, good or bad, morally permissible or forbidden, etc. is entirely immaterial to its status as a *promoter* of helping and sharing behavior or as a *deterrent* of cheating, defection, or any other uncooperative behavior. The fact that moral belief may play such a causal role is merely to attribute to it certain behavioral consequences, not any specific content. Not only do we not know that moral belief is about human prosocial affairs, we can scarcely make sense of the idea that they could be about such a thing. We do not seem to be in a position to know that moral belief is somehow about prosociality.

Finally, we also don't know whether there is some sort of reliability-guaranteeing psycho-moral nexus. If such a nexus did exist of either the constructivist or externalist kind, then no human could be completely unreliable about moral matters, since there would be a metaphysical guarantee of at least some overlap with the truth. But it is possible for fallible creatures like us to be completely wrong about morality. Presumably, there's no limit the atrocities we might find to be morally permissible or right in our darkest dreams. Since knowing such a nexus exists involves rejecting the very possibility of radical moral error, insofar as we regard this possibility as genuine, we are not in a position to know there is any kind of psycho-moral nexus. In any case, even if one did not consider this possibility to be genuine, that's a far cry from ruling in the existence of such a nexus as actual. Much more would be needed to establish such a thing as fact.

Now, since we are not in a position to know any one conjunct of the moral ignorance hypothesis is false, we are not in a position to know this triple conjunction is false and, by the closure of knowledge under competent deduction, we do not know whether there is such a thing as human moral knowledge. Following a formulation from Hawthorne (2013), the underlying closure principle is:

> MULTI-PREMISE DEDUCTIVE CLOSURE (MDC): If one knows some premises and competently deduces q from those premises, thereby coming to believe q, while retaining one's knowledge of those premises throughout, one comes to know q.

If we know there is human moral knowledge, and we know that if there is the moral ignorance hypothesis is false, then we may competently deduce and thereby come to know its falsity. But since we aren't in a position to know this, it would seem that we don't really know there is moral knowledge, after all. Given MDC, then, we don't know whether humans know right from wrong, good from evil, or how humans morally ought, or ought not, to act. Assuming we have never been in a position to know either that moral cognition did not descend truth-orthogonally or that there is a psycho-moral nexus, we have never been in a position to know that human moral knowledge exists.

Given Darwinian naturalism, then, since before Darwin we lacked the conceptual resources to be in a position to know that there is human moral knowledge and, thereafter, we have been unable to secure the empirical evidence to know such knowledge exists, contrary to what the long history of human morality suggests, we humans have never known there is such a thing as human moral knowledge. In the following sections, I address a variety of objections.

2.3.2 The autonomy response

Many are inclined to think that evolutionary hypotheses like the moral prosociality hypothesis concern biological matters of no ethical relevance. Specifically, some think that the moral prosociality hypothesis is just a piece of adaptationist speculation and that, as such, it can be dismissed as irrelevant to our claim to moral knowledge. The thought is that, just as the *mere* possibility of demonic deception fails to establish our ignorance of the existence of perceptual knowledge, the mere possibility of the moral

ignorance hypothesis fails to cast any credible doubt on our knowledge of human moral knowledge because the moral *prosociality* hypothesis does not merit serious consideration.

But the moral ignorance hypothesis cannot for this reason be dismissed as irrelevant or otherwise negligible because the moral prosociality hypothesis is indeed a *live scientific hypothesis*. Not only does it merit serious consideration, it routinely gets it. Since *moreover* we don't know this hypothesis is *not* truth-orthogonal (since we don't know moral beliefs *are* about prosocial matters), and *furthermore* we don't know there is a constitutive psycho-moral connection, we seem not to know human moral knowledge exists. We should take the moral ignorance hypothesis seriously precisely because the evidence in favor of the moral prosociality hypothesis is so clear that its exact import is hotly debated.

Along similar lines, some ethicists have claimed ethical inquiry is otherwise autonomous from the findings of evolutionary biology. Nagel (1979) argues that the domain of ethics is normatively isolated from evolutionary biology because it has its own "internal standards of justification and criticism" and that it is therefore immune to challenges from biological considerations (p. 42). He presumably no longer thinks this, since in *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False* he finds Street's (2006) Darwinian Dilemma so pressing as to warrant the strongly worded subtitle. But others continue to find this suggestion quite compelling.

FitzPatrick (2008) defines the Nagelian sense of autonomy as involving "exercises of thought that are not themselves significantly shaped by specific evolutionarily given tendencies, but instead follow independent norms appropriate to the pursuits in question" (2008: §2.4). He then writes that:

> [The assumption of Nagelian autonomy] seems hard to deny in the face of such abstract pursuits as algebraic topology, quantum field theory, population biology, modal metaphysics, or twelve-tone musical composition, all of which seem transparently to involve precisely such *autonomous applications of human intelligence*. Even if there are evolutionary influences behind our general tendency to engage in certain kinds of mental activity [...], this would not show that the activity is governed in its details by such influences. (ibid.; emphasis in original)

FitzPatrick thinks that for an evolutionary challenge to have any force evolutionary influences must govern mental activity *in its details*. But nobody (I hope!) thinks that that's genuinely possible. In any case, Nagelian autonomy is orthogonal to the evolutionary challenge from ignorance. The issue is whether we know *living, breathing* ethicists can learn any moral truths, not whether ethics has "internal" epistemic standards. Even if it did have these, had moral cognition descended truth-orthogonally, in the absence of a psycho-moral nexus no ethical conclusion *could* amount to knowledge because moral belief, if true, would be accidentally true. FitzPatrick's appeal to autonomy works only against the crudest of strawmen. More generally, to avoid the challenge, it is not normative ethics that must be "autonomous" but human ethicists that must be free of Darwinian origins, for it is the presence of the human capacity for moral knowledge that is in question, not the supposed normative isolation of the domain of ethics.

2.3.3 The third-factor response

A popular response to Street's "Darwinian Dilemma" is the so-called third-factor response. Some might suspect that it can be used to respond to this challenge. On third-factor accounts,[40] a single factor helps causally make it the case that we tend to have the moral beliefs we do, as well as help metaphysically make it the case that these beliefs' content is true. (Cf. Berker 2014.) The idea is that such a factor is among the debunker's so-called "non-truth-tracking evolutionary forces" and that therefore an appeal to these forces in the ultimate explanation of our moral beliefs does not show that our having many true moral beliefs is a fluke, even given moral realism. This response, however, does not do anything to defend against the present challenge. The cogency of the evolutionary challenge from ignorance does not hinge on the claim that evolution *need not* have made our having many true moral beliefs a fluke. The central epistemological issue is, rather, whether we know moral cognition descended truth-dependently or that it has some other nonaccidental connection to moral truth. In reply to this challenge, the third-factor response would be a complete non-sequitur, since at no point does it question whether moral cognition could have had a truth-dependent descent.

---

[40] Nozick (1981), Copp (2008), Enoch (2010), Wielenberg (2010), Brosnan (2011), Skarsaune (2011), and Berker (2014) all provide examples.

2.3.4 The overgeneration response

A more serious objection is that the challenge overgeneralizes. If the challenge were to target instead our claim to mathematical knowledge, for instance, would we be saddled with the conclusion that we don't know whether there is human mathematical knowledge? If we are forced to conclude this, this thought runs, then perhaps the challenge need not be taken so seriously, for clearly something must be terribly wrong with it. Surely, we know that we humans know that $1 + 1 = 2$.

Fortunately, while we may not be able to rule in a psycho-mathematical nexus, we can rule out mathematical cognition's truth-orthogonal descent. Suppose we create a model of a bridge. To test the accuracy of the model, we calculate the least amount of stress needed to collapse it and then subject it to it. Lo and behold, the bridge begins to break down at precisely the indicated stress level. Since we can see the bridge crashing down, and we realize that it began to do this at the predicted level of stress, we get confirmation of the accuracy of the model. Since the seeing and the realizing does not involve any mathematical cognition, the accuracy of the confirmation is independent of mathematical cognition's reliability. The prediction's success is therefore evidence for the accuracy of the mathematical calculations that went into the construction of the model. Since mathematical calculations frequently lead to correct predictions in this way, we have excellent evidence against the truth-orthogonal descent of mathematical cognition.

Importantly, I am not here *assuming* that mathematical entities are relevant to the causal structure of the world and so that certain forms of mathematical realism are false. Rather, the fact that mathematical calculations frequently support successful prediction is evidence for the causal relevance of mathematical entities *and* against the truth-orthogonal descent of mathematical cognition. If views in the philosophy of mathematics make the success of practical applications puzzling, that's a mark against the views, not a reason to be skeptical of the success. This is a large part of why the central epistemological problem of mathematics has been to *explain* our reliability on mathematical matters, not to justify the very idea that we are reliable (Benacerraf, 1973; see also Field 1989).

Our claim to mathematical knowledge faces the evolutionary challenge from ignorance, but we can dispense with it relatively easily. Rather than foolishly generalize, we more clearly see just how much

more difficult it is to sustain the claim to human moral knowledge in the face of the challenge. It can be met in the mathematical case despite the supposed causal impotence of the target facts. This reinforces the earlier point that the obstacle to knowledge is not in the nature of the target facts but the nature of the putative knower. The challenge's message that we don't know enough about ourselves to know whether it is in our nature to know moral truths is, if anything, made to ring more notably true.

Some may still worry that the challenge overgenerates in a different direction. If it can target our claim to perceptual knowledge, the challenge would generalize globally, since we wouldn't be able to rely on empirical evidence to rule out the truth-orthogonal descent of human cognition generally. Furthermore, since we cannot rely on any empirical evidence, we cannot rationally motivate the idea of a reliability-guaranteeing link between human cognition and its subject matter because we cannot take our reliability about anything to be in evidence. We seem to be led to conclude, then, that we do not know there is *any* human knowledge. If our claim to perceptual knowledge cannot fend off the evolutionary challenge, it may be run as a *reductio* of its premises, or so this worry goes.

This reply supposes there is human epistemological knowledge. But if *ex hypothesi* we cannot rule out the truth-orthogonal descent of perceptual cognition, then we cannot justify our claim to epistemological knowledge, since its truth-orthogonal descent we would not be able to rule out and the appropriate nexus we would not be able to motivate. These things would require *some* empirical evidence. Since the challenge's premises refer to knowledge, and we cannot rely on epistemological knowledge, we could not know the premises well enough to derive any justification from them. Without the knowledge that accidentally true belief cannot amount to knowledge or that knowledge is closed under competent deduction, we cannot raise the challenge to *any* skeptical effect. The challenge cannot be run as a *reductio* of its premises.

This last observation also takes care of the worry that the evolutionary challenge from ignorance might merely be an instance of a more general epistemological problem. Berker (2014) has argued that Street's Darwinian Dilemma may only be raising the problem of "how to justify our reliance on our most basic cognitive faculties without relying on those same faculties in a question-begging manner" (p. 215). Vavova (2014), for another example, has suggested that the evolutionary challenge may just be the

challenge of "justify[ing] your entire body of belief [...] without presupposing the truth of any of the beliefs that have been called into question" (p. 93). But because the evolutionary challenge from ignorance is not *capable* of targeting our claim to perceptual knowledge, it cannot be an instance of either of these because it can't *possibly* match either of these problems in scope.

Mathematical knowledge isn't terribly threatened by the challenge, and perceptual knowledge *can't* be one of its casualties. Unfortunately for our claim to moral knowledge, we have no equally good reason to think that humans evolved the means to learn any moral truths. The evolutionary challenge from ignorance does not foolishly generalize. In the final section, I address the thought that our ignorance of the existence of human moral knowledge is any less worrisome than our being morally ignorant.

2.3.5 The "So What?" reaction

Some might take comfort in the fact that the evolutionary challenge from ignorance suggests only that we do not know there is moral knowledge. This reaction would be a mistake. Even though it is our knowledge of human moral knowledge that is in doubt, not its existence, the challenge is just as worrying as any direct rebuttal of the claim that there is human moral knowledge because they are dialectically on a par. On the challenge from ignorance, we don't know there is moral knowledge because we don't have the evidence to know that the means to acquire moral knowledge ever evolved. Direct rebuttals present reasons for thinking that we do *not* have moral knowledge and so that we are not justified in thinking that we have it. Either way, the putative knower faces the challenge of justifying their belief in the existence of moral knowledge. Nothing peculiar to the evolutionary challenge makes it any less troubling than attempts to show that we have no moral knowledge at all. If direct rebuttals trouble you, then so should the evolutionary challenge from ignorance.

2.4 Conclusion

I distinguished between truth-dependent and truth-orthogonal selection and argued that the moral prosociality hypothesis may be both truth-orthogonal and biologically possible. I then argued that it is incompatible with the existence of human moral knowledge, and noted that unless we can rule out the truth-orthogonal descent of moral cognition, or rule in a human psycho-moral nexus, we don't know

if we evolved the capacity to acquire moral knowledge. I then explained that since we don't have enough evidence against the moral ignorance hypothesis, if we are to accept both the principle of deductive closure and the doctrine of Darwinian naturalism, we must forgo our claim to moral knowledge.

But if all of this is right, then why resist the skeptical solution to the problem of moral knowledge? If, empirically, we cannot tell whether there is moral knowledge, then we ought to view moral skepticism as equally deserving of exploration or defense as any anti-skeptical response because our gut feeling that we do have this knowledge would just be an embodiment of a preconceived opinion that is not based on our understanding of the human mind or its evolutionary origins. If nothing we know about the origin of our biology suggests that, indeed, we can tell right from wrong, then moral skepticism's patina of implausibility should not be taken to be suggestive of its truth-value, for if we do have this ability it evolved and if it evolved its existence is nothing more than a biological fact.

As a Darwinian naturalist and believer in the fruitfulness of deduction, I think that we can only conclude that we do not know whether we know good from evil, right from wrong, or how we morally ought, or ought not, to act. I believe that it is time we take the skeptical solution to the problem of moral knowledge much more seriously.

References

Benacerraf, P. (1973). Mathematical truth. *The Journal of Philosophy*, 661-679.

Berker, S. (2014). Does evolutionary psychology show that normativity is mind-dependent? In J. D'Arms, & D. Jacobson (Eds.), *Moral psychology and human agency: Essays in the new science of ethics.* Oxford: Oxford University Press.

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 171-195.

Brasier, M. D., Norman, D. B., Liu, A. G., Cotton, L. J., Hiscocks, J. E., Garwood, R. J., . . . Wacey, D. (2016). Remarkable preservation of brain tissues in an Early Cretaceous iguanodontian dinosaur. *Earth System Evolution and Early Life: A Celebration of the Work of Martin Brasier, 448.* doi:10.1144/SP448.3

Brosnan, K. (2011). Do the evolutionary origins of our moral beliefs undermine moral knowledge? *Biology and Philosophy*, 51-64.

Copp, D. (2008). Darwinian skepticism about moral realism. *Philosophical Issues*, 186-206.

Darwin, C. (1859/2009). *The origin of species by means of natural selection the preservation of favoured races in the struggle for life* (150th anniversary ed.). New York: Signet Classics.

Enoch, D. (2010). The epistemological challenge to metanormative realism: How best to understand it, and how to cope with it. *Philosophical Studies*, 413-438.

Field, H. (1989). *Realism, mathematics and modality.* Oxford: Blackwell.

FitzPatrick, W. (2008, December 19). *Morality and evolutionary biology.* Retrieved from Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/entries/morality-biology/

Fraser, B. J. (2014). Evolutionary debunking arguments and the reliability of moral cognition. *Philosophical Studies*, 457-473.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 121-123.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, 1-16.

Haselton, M. G. (2003). The sexual overperception bias: Evidence of a systematic bias in men from a survey of naturally occurring events. *Journal of Research in Personality*, 34-47.

Hawthorne, J. (2013). The case for closure. In M. Steup, J. Turri, & E. Sosa (Eds.), *Contemporary debates in epistemology* (2nd ed., pp. 26-42). Wiley Blackwell.

Johnson, D. (2015). *God is watching you: How the fear of God makes us human.* New York: Oxford University Press.

Joyce, R. (2001). *The myth of morality*. Cambridge: Cambridge University Press.

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

Joyce, R. (2016). Evolution, truth-tracking, and moral skepticism. In B. Reichardt (Ed.), *Essays in moral skepticism* (pp. 142-158). Oxford: Oxford University Press.

Kahane, G. (2011). Evolutionary debunking arguments. *Nous*, 103-125.

Lewontin, R. C. (1998). The evolution of cognition: Questions we will never answer. In *An invitation to cognitive science: Methods, models, and conceptual issues* (Vol. IV, pp. 107-132). Cambridge, MA: The MIT Press.

McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, 493-510.

Miller, G. F. (2007). Sexual selection for moral virtues. *The Quarterly Review of Biology*, 97-125.

Nagel, T. (1979). Ethics without biology. In *Moral questions* (pp. 142-146). Cambridge: Cambridge University Press.

Nagel, T. (2012). *Mind and cosmos: Why the materialist Neo-Darwinian conception of nature is almost certainly false*. Oxford: Oxford University Press.

Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Harvard University Press.

Pritchard, D. (2007). *Epistemic luck*. Oxford: Oxford University Press.

Richardson, R. C. (2007). *Evolutionary psychology as maladapted psychology*. Cambridge, MA: MIT Press.

Ruse, M. (1986). *Taking Darwin seriously: A naturalistic approach to philosophy*. Oxford: Basil Blackwell.

Setiya, K. (2012). *Knowing right from wrong*. Oxford: Oxford University Press.

Skarsaune, K. O. (2011). Darwin and moral realism: Survival of the iffiest. *Philosophical Studies*, 229-243.

Smithsonian Institute. (2016, December 30). *Human Fossils*. Retrieved from The Smithsonian Institute's Human Origins Program: http://humanorigins.si.edu/evidence/human-fossils

Sober, E. (1984). *The nature of selection: Evolutionary theory in philosophical focus*. Chicago: University of Chicago Press.

Sober, E. (2006). Models of cultural evolution. In E. Sober (Ed.), *Conceptual issues in evolutionary biology* (pp. 535-551). Cambridge, MA: MIT Press.

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution of psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

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

Street, S. (2008). Constructivism about reasons. *Oxford Studies in Metaethics, III*, 207-245.

Tomasello, M. (2016). *A natural history of human morality*. Cambridge, MA: Harvard University Press.

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

Vavova, E. D. (2014). Debunking evolutionary debunking. *Oxford Studies in Metaethics, IX*, 76-101.

White, R. (2010). "You just believe that because ...". *Philosophical Perspectives*, 573-615.

Wielenberg, E. (2010). On the evolutionary debunking of morality. *Ethics*, 441-464.

CHAPTER 3

Human Morality: Lie or Heirloom?

Lies can be told about anything and anything told to you might be a lie. But some lies are more consequential than others. Some, as I will argue here, give rise to tremendous phantoms of misbelief. In this chapter, I argue that morality might have begun as a lie told by our distant ancestors, a lie that turned out to be compelling or adaptive enough to spread across the population, becoming culturally elaborated enough to generate morality as we know it. Call this the *morality as lie* hypothesis (*the lie hypothesis* for short). This suggestion is in stark contrast to what I will call the *morality as heirloom* assumption.

Shared by many contemporary philosophers, including some at extremes of metaethical and methodological spectrums, the morality as heirloom assumption takes the descent of moral belief to have involved the epistemological stewardship of our forebears: our parents taught us to tell right from wrong, their parents taught them, and so on to the infancy of morality. Here, for example, is Gideon Rosen:

> I myself favor a view according to which the moral facts are both utterly objective and perfectly inert. How might we know them if that's what they're like? Well, they might be requirements of pure practical reason, in which case we might know them by noting that their denials involve us in "practical contradictions". But much more plausibly, it might simply be that we learn moral principles from our parents which we then refine and revise in light of experience and reflection according to a critical discipline which we also inherit. (1998, p. 398)

And Philip Kitcher (2012, p. 1):

> Ethics is something human beings have been working out together for most of our history as a species. The needs that prompt the cooperative activity of the ethical project lie deep in our human characteristics, and were focused sharply in our human past. Over tens of thousands of

> years, different human societies have conducted "experiments of living", in Mill's apt phrase, trying to find ways of attending to the difficulties inherent in a form of social life to which evolution inclined our pre-human ancestors.

I aim to show that this widely held view of our moral heritage as having involved a kind of ethical curation fails to properly explain the genesis of moral belief and cannot be plausibly amended to explain it. A much better if troubling hypothesis contends that moral opinion first came about as the result of ancient parental deception which was elaborated over tens of thousands of years by the increasingly sophisticated, and eventually unwitting, rumor-mongering of our ancestors. To be sure, there won't be an attempt to mount a full-scale defense of this lie hypothesis in what follows; my more modest intention is to outline and defend an evolutionary perspective that is scientifically plausible and that promises to fill this gap in our current understanding of moral evolution.

This chapter is organized as follows. After providing some background information on the moral evolution literature, I sketch the lie hypothesis, clarify its content, defend it from objections, and show how it explains humans' first moral beliefs as the result of ancient parental deception (§3.1). I then attempt to develop the morality as heirloom assumption into a hypothesis that can explain how moral belief first arose and find that on most of the possible origin scenarios the heirloom hypothesis cannot explain how the first moral belief came about (§3.2). I argue to end that the lie hypothesis is explanatorily superior to the heirloom hypothesis and that to the best of our knowledge morality is a lie (§3.3).

3.1 The Lie Hypothesis and the Advent of Moral Belief

Moral evolution theories have focused on just two elements of moral evolution: the biological evolution of the psychological capacity for moral cognition and the cultural evolution of the proliferation and differentiation of moral beliefs.[41] In this section, I discuss instead what I will call *the advent of moral belief*, the sequence of events that led to the formation of the *first* moral beliefs. The advent of moral belief

---

[41] Most moral evolution theories address both aspects. See, e. g., Alexander (1987), Sober and Wilson (1999), Joyce (2006), Hauser (2006), de Waal (2009), Kitcher (2011) and Tomasello (2016). For a smattering of the moral evolution literature, see *Behaviour*'s Special Issue on evolved morality (de Waal, Churchland, Pievani, & Parmigiani, 2014).

lies dab smack in the middle of the other two more widely discussed aspects of moral evolution. Before moral belief can spread, it must be held, and it can only be held once the capacity to attribute moral properties is present. The advent of moral belief is the actual historical development that led our ancestors to *believe* (as opposed to suspect or wonder whether) people, actions, or events have moral properties.

No corner of the moral evolution literature known to me touches on the advent of moral belief. Obviously, it occurred, but since it probably only happened once and, at the very least, tens of thousands of years ago, it is not a topic about which we can hope to ever tell the actual story. That's probably why it is so neglected despite its obvious importance to our understanding of the evolution of morality. In any case, we can theorize responsibly about what might have transpired at that time by aiming only to tell a story that is consistent with the available empirical evidence and with background constraints (Kitcher, 2011). In this section, I sketch the contours of such a "how possibly?" explanation, specifically, I outline and defend a hypothesis that appeals to ancient parental deception to explain how moral beliefs might have first been formed. But first, a note on the background constraints.

3.1.1 The byproduct approach and the moral prosociality hypothesis

The moral evolution literature is home to many controversies, but there is a striking amount of agreement on two important points. First, though it might not seem that way at first glance, it is widely assumed that the capacity for moral cognition *first* emerged as a byproduct of cognitive mechanisms with their own separate, distinct evolutionary rationales. It might not seem that way initially because there is a great deal of disagreement, first, on whether the capacity has been adaptive since then and, second, if it has, whether natural selection has operated on it for long enough to become adapted and so by now be an adaptation (see Joyce 2014).[42] Evolutionary biologist Francisco J. Ayala argues, for instance, that "ethical behavior came about in evolution not because it is adaptive but as a necessary consequence of

---

[42] Evolutionary biologists Gould and Vrba (1982) coin the term *exaptation* to refer to traits that after their emergence as byproducts of natural selection become co-opted to perform an adaptive role at a later evolutionary stage of a population and may in turn over subsequent generations become an adapted feature of its members.

As Gould (1991) notes, the construct of an exaptation is crucial to developing a realistic and practicable scientific discipline of evolutionary psychology. There are also general theoretical reasons for expecting the use of the byproduct approach.

man's eminent intellectual abilities, which are an attribute directly promoted by natural selection" (2010, p. 9015). Richard Joyce, by contrast, advocates the thesis that "human morality is a distinct adaptation wrought by biological natural selection" (2008, p. 213).

While these two men disagree on whether moral cognition is *presently* an adaptation, both agree that the capacity for moral cognition, i.e., the ability to make moral judgments, was initially an effect of psychological mechanisms that evolved to perform tasks of nonmoral cognition. These enabling mechanisms presumably include the capacities for sympathetic concern, self-regulation and belief formation, and the abilities to discriminate, abstract from, and classify persons, events and actions and to anticipate and predict consequences of one's own actions and that of one's fellows (Kitcher, 2006; Deem, 2016). Ayala takes the ability to anticipate the consequences of one's actions specifically with respect to other people to have been critical to the emergence of the ability to evaluate actions morally; whereas Joyce (2006) and his followers, e.g., James (2011), take the development of early moral tendencies to involve the adaptive, if excessive, use of the ability to classify members of one's close-knit group as kin, regardless of their actual degree of genetic relatedness.

Even former Harvard psychologist Marc Hauser, who posits an evolutionary advance in which our ancestors acquire a "moral organ," proposes, centrally, that "one branch of the root of our moral judgments can be found in the nature of expectations concerning actions" (2006, p. 168). Most moral evolution theorists take a byproduct approach to the problem of explaining the initial biological emergence of the capacity for moral cognition.

The second point of broad scientific agreement is what I call *the moral prosociality hypothesis*: roughly, that moral cognition evolved to help enable the evolution of modern human cooperation by promoting sharing and helping behavior and/or deterring cheating and defection among non-kin. Central among Joyce's (2006) evolutionary hypotheses is, for instance, the suggestion that the tendency to direct helping behavior towards non-kin came about as our early human ancestors' prosocial bias towards kin expanded to include non-kin. Ayala (2010), to continue with our two foes, suggests – echoing Darwin in *The Descent of Man* – that, because humans can understand the social benefits of altruistic behavior for the group and so indirectly for the individual, this maladaptive but prosocial practice can spread across

the population despite the cost incurred at the level of the individual organism. Most moral evolution theorists accept the moral prosociality hypothesis as a starting point for understanding the proliferation and differentiation of moral beliefs.

In line with the byproduct approach and the moral prosociality hypothesis, in developing the lie hypothesis I will assume the capacity for moral cognition originated as a byproduct of evolved cognitive mechanisms and that after the advent of moral belief it was primarily their relative prosociality that led to their differential dissemination across the broader population.

3.1.2 A brief history of humanity and the advent of moral belief

A sense of certain milestones in the timeline of human evolution will be useful in what follows. Recorded history began about 6,000 years ago (Houston, 2004), but the modern human species is much, much older. Genetic evidence from early modern humans suggests that our species is approximately 200,000 years old (Trinkaus, 2005). The Upper Paleolithic Revolution, an unprecedented cultural explosion that resulted in, among many other things, artwork, advanced stone tool technology, and complex rituals and social structures, is estimated to have begun about 40,000 years ago (Bar-Yosef, 2002). While there is no consensus view,[43] the origin of language is usually set sometime between these last two estimates because language is frequently thought to be, uniquely, a human innovation and to have been present throughout the Upper Paleolithic Revolution.[44]

Human language is thought to have been present throughout this Revolution because full language abilities are supposed to have been required for the advancements of the time. The development of pictorial representations, such as painted and engraved imagery, strongly suggests the capacity for symbolic thought. Likewise, the development of complex ritual systems and social structures, together with increases in social group size (Richerson, 2013), suggests the presence of non-kin cooperation and by the moral prosociality hypothesis that of shared moral belief systems. Philip Kitcher suggests as a

---

[43] A matter of controversy, one of many, is the degree to which our hominid progenitors and their contemporaries, like the Neanderthals (D'Anastasio, et al., 2013), had human-like language. For a sobering survey of the empirical obstacles to resolving the mystery of language evolution, see Hauser, et al. (2014).

[44] For a comprehensive survey of the language evolution literature, see *The Oxford Handbook of Language Evolution*.

conservative estimate that the first ventures into ethical practice occurred fifty thousand years ago and "probably involved group discussion, on terms of rough equality, directed towards issues of sharing and intragroup aggression" (2011, p. 11). Following Kitcher, I will assume the advent of moral belief occurred sometime between the emergence of language and the beginning of the Revolution and so at the very least 40,000 years ago.

While the suggestion that language predates moral belief systems is not entirely uncontroversial, it is a reasonably safe assumption. Although adaptationists, such as James (2011), often take morality to be significantly older than language, they think moral evolution begins with the development of prosocial biases towards non-kin, not with the advent of moral belief or with the emergence of full-fledged moral cognition. I doubt any of them think shared moral belief systems predate the emergence of language. But to make this assumption even safer, unlike Kitcher , I will not assume the advent of moral belief occurred after human beings acquired *full* linguistic abilities, which is taken to have occurred at most 100,000 years ago, with the development of the modern vocal tract (Enard, et al., 2002). For present purposes, the key characteristic of human language will be its *productivity*: the capacity to communicate an infinite number of ideas using a limited set of words (Chomsky, 1957; 1969).[45] So, I will assume that at the earliest the advent of moral belief occurred after the evolution of productive language, which – seeing as productivity is quite probably a feature unique to human systems of communication (Hauser, Chomsky, & Fitch, 2002) – occurred at most 200,000 years ago.

---

[45] Later in the text, when most relevant, I discuss the productivity of language in greater detail. Productivity is commonly assumed to be a universal feature of human language (but see Everett 2005 for evidence suggesting that the Pirahã, an Amazonian tribe, have a finite language).

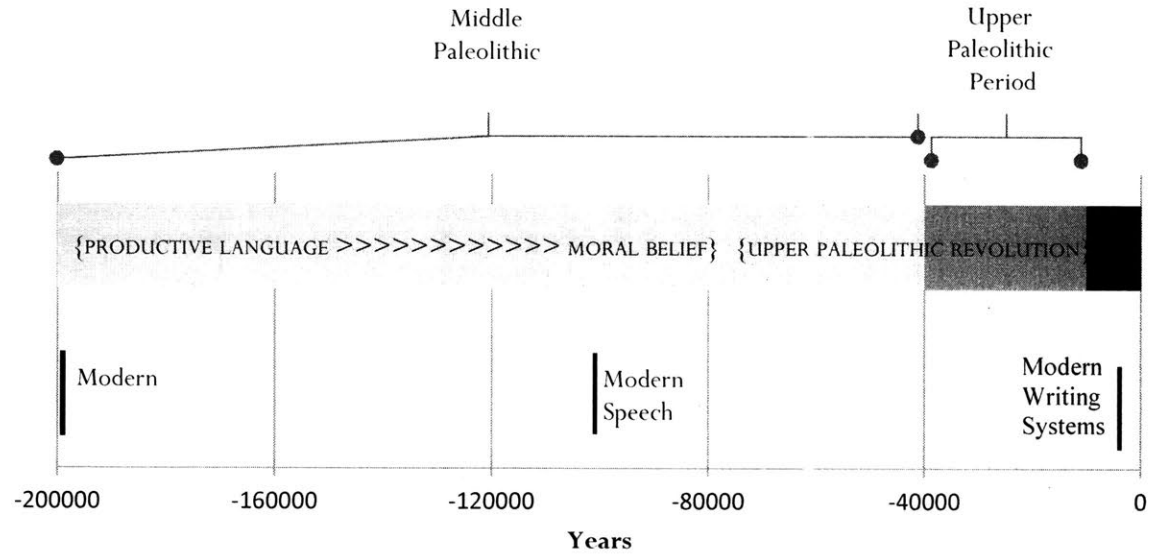

Fig. 4. A brief history of humanity. The origin of productive language precedes the advent of moral belief. The Upper Paleolithic Revolution is the transition from the Middle Paleolithic to the Upper Paleolithic Periods.

I assume in what follows that the advent of moral belief occurred *after* the evolution of productive language, at most 200,000 years ago, and *before* the Upper Paleolithic Revolution, at the very least 40,000 years ago. Now, onto the hypothesis.

3.1.3 THE MORALITY AS LIE HYPOTHESIS

What if one day tens of thousands of years ago, sensing that their innocence or trust was such that they would believe her, knowing that, if they did believe her, they would behave, a mother tells her children that certain actions are wrong no matter what or right come what may? By her lights, there is no such action, since an action is right or wrong only as means to an end, but like all human beings of the time, her children included, she can creatively combine old ideas to generate new ones and old words to express novel thoughts. She can form these novel thoughts, moreover, without believing their content (or, if she must initially believe it, she does not believe it for long). [46]

---

[46] Here I allude to the two psychological models of the comprehension and acceptance of propositions (Gilbert, 1991). On the Cartesian model, comprehension precedes and is separate from acceptance: one must comprehend a proposition before one can believe it. On the Spinozian model, acceptance is identified with comprehension and distinguished from truth-evaluation: one must believe a proposition to comprehend it and only then can its truth be

Her folk-psychological understanding of her fellow human beings is advanced enough for her to know that talk triggers belief and belief causes behavior. Specifically, she knows even if not under this description that utterances[47] *can trigger belief with their content and that it is in virtue of its content that a belief causes the behavior it does. She understands in her gut that a good way to get her children to cooperate is to inculcate firmly held beliefs about the wrongness or rightness of certain key acts.*

The children, trusting their mother implicitly, believe her. They possess all her same cognitive abilities, even if some only in a less advanced form, for all members of the species had these abilities as part of their species-typical cognitive endowment. Specifically, like their mother's language, theirs is productive, allowing them to both produce and understand utterly novel utterances. Their mother differs from her contemporaries in that she is the first to make use of her cognitive and linguistic abilities to hatch and execute this specific plan. Hers is an evolutionarily novel application of the human capabilities of the time.

Once presented with it, her peers see the brilliance of her parental strategy, specifically, its potential to save them a great deal of effort and time. So, parents in her group adopt the practice and, over subsequent generations, it spreads across the entire human population. Over the same period, the strategy is refined and revised in light of parents' experience and reflection, improving the execution of the core idea and developing a knack for determining which kinds of action can be usefully said and credibly told to be unconditionally right or wrong.

In time, the mother's invention of this parental strategy is forgotten. Children, initially fooled into thinking that there is an absolute right or wrong, come to be sincerely believe this and end up teaching it to their own children as known fact. By the time of the earliest extant records, about 6,000 years ago, the practice had evolved to become the complex forms of ethical life recognizable in the world of that time. Today we continue to live out the repercussions of this most formidable lie.

---

critically assessed. The advent of moral belief amounts to more than the formation of the first moral beliefs on the Spinozian model, since that may merely involve their initial contemplation. I go on this brief excursion because, believe it or not, there is compelling empirical evidence for the Spinozian model. For its philosophically informed discussion, see Mandelbaum (2014).

[47] With the term *utterance*, I refer to both spoken and signed linguistic expressions.

This is, of course, a very rough sketch of the lie hypothesis, but we can already note an important point. Since the lie hypothesis is explicit about its consistency with the historical record, and neither the fossil nor the prehistorical archeological record contains information for or against the occurrence of such specific sequences of events, the lie hypothesis is consistent with the empirical evidence available on the advent of moral belief, since the historical, archeological, and fossil records exhaust the evidence available on human evolution between 200,000 and 40,000 years ago. Given its consistency with the empirical constraints, to establish its status as a "how possibly?" explanation, I must show that the lie hypothesis is consistent with the above background constraints.

Accordingly, in the coming sections, I draw on scientific and empirical detail to show how the hypothesized sequence of events is permitted by the byproduct approach and the moral prosociality hypothesis. First, I explain why the hypothesized suite of cognitive abilities can be plausibly assumed to be present by the advent of moral belief (§1.4). Second, I give an account of how the mother might have first formed moral thoughts (§1.5). Third, I explain why it is plausible that her testimony would have caused lasting belief in her children (§1.6). Fourth, relying on a model by selection theorists Joseph Henrich and Robert Boyd, I propose an evolutionary mechanism by which the mother's parental strategy might have spread to the rest of the population (§1.7). Lastly, I explain how the parental strategy might have become applied beyond children to adult non-kin peers and in consequence how the lie hypothesis can be thought to merge with the moral prosociality hypothesis (§1.8).

But before proceeding, I want to make two clarificatory notes. First, the lie hypothesis is not committed to the existence of such a brilliant individual as the mother. She is merely a personification of the set of cognitive achievements responsible for the advent of moral belief. As a reminder of this, I shall call her *the Mother*. Perhaps whoever first communicated the idea of an absolute right and wrong differs from the individual who thought of its application in parenting, that person in turn differs from the first person to trigger moral belief in children, that person, moreover, may differ from the first person to trigger lasting moral beliefs in their children, and so on. Second, these individuals are unaware of having reached their respective cognitive milestone. The Mother need not have an advanced understanding of the fact that she believes actions are right or wrong only to the degree that these are fit to achieve the desired ends, nor must she have a nuanced appreciation of the benefits of her parenting technique. She

must only have a gut feeling or intimation that suffices to reflect the presence of its respective cognitive achievement. Self-understanding of a more sophisticated type would presumably be absent, since these achievements would have then been too recent in human evolution to plausibly suppose anyone would have developed the concepts to grasp their occurrence, least of all without the technology of writing.

3.1.4 The capacity for moral cognition

In addition to productive language, the lie hypothesis assumes human cognition is developed enough for the Mother to execute her brilliant idea. This assumption presupposes that (i) adult human cognition was advanced enough to form the idea behind the Mothers' lies and that (ii) the children were cognitively sophisticated enough to understand and believe their content. Two theories of the evolution of human cognition are helpful in thinking about possible evolutionary pathways to this suite of cognitive abilities. Peter Godfrey-Smith (1996) argues that the biological function of complex cognitive abilities is to enable an organism to deal with complex environments, for instance by enabling the organism to adjust to complexly changing circumstances and to react adaptively to ecological variability. Kim Sterelny (2003) argues that complex representational systems would be the target of natural selection in populations within heterogenous and informationally opaque environments, leading to the evolutionary development of capacities for decoupling representational states from reflexive, cue-bound behavior, as in counterfactual thinking and means-end reasoning.

On both theories, the demand for complex cognitive faculties would have been especially keen in the context of the social and cultural environment of our ancestors (Dunbar, 1998; Sterelny, 2012). These complex cognitive faculties are precisely the enabling mechanisms already widely assumed to be responsible for the presence of the capacity for moral cognition. By the lights of both theories of the evolution of complex cognition, then, the lie hypothesis does not posit cognitive resources that go beyond those allowed for by the byproduct approach, such as, to repeat, the capacities for sympathetic concern, self-regulation and belief formation, and the abilities to discriminate, abstract from, and classify persons, events and actions, and to anticipate and predict consequences of one's own actions and that of one's fellow humans.

The sequence of events described by the lie hypothesis is permitted by recent theories of the evolution of complex cognition and moral evolution theories that adopt the byproduct approach. In the next section, I explain why the Mother has the capacity to form moral thoughts given the presumed suite of cognitive capacities.

3.1.5 The generation and communication of moral thoughts

The lie hypothesis assumes that the Mother has productive language. As someone with productive language, she can combine pre-existing linguistic expressions to generate utterly novel and meaningful expressions and she can understand these expressions despite their utter novelty. Since she can therefore register and exploit the compositional nature of linguistic representation in this way, she must have the cognitive capacity to creatively combine pre-existing mental representations both to generate and to understand utterly novel, well-formed thoughts. Given this capacity, to form the thought that a certain action is unconditionally right or wrong, she need only have the constituent representations.

Furthermore, on the lie hypothesis, the Mother has these representations on account of her abilities to discriminate, abstract from, and classify actions, and to reason instrumentally and modally. Specifically, since she reasons instrumentally, she has a conception of the rightness or wrongness of actions as means to an end and, since she reasons in modal terms, she has at her disposal the concepts of possibility and necessity. Therefore, she can consider whether, and so conceptualize the possibility that, an action may be right or wrong no matter what, yielding either the thought that the target action is unconditionally right or that it is unconditionally wrong.

It is the Mother's unprecedented and suitably coordinated exercise of certain of her cognitive abilities that would lead her to first form moral thoughts. If she hasn't yet the linguistic expressions for the constituent representations are available, she would only be able to communicate these thoughts once these are at hand. If these expressions predate the moral thoughts, then they may have served as aids in the formation of her first moral thoughts. The lie hypothesis is silent on which of these two origin scenarios is correct.

In either case, she expresses those moral thoughts that she judges capable of eliciting a lasting and behavior-informing belief in her children. In making these folk-psychological judgments, she need only rely on her capacity for sympathetic concern and her ability to anticipate and predict consequences of her own actions and that of others' actions, for these require a sophisticated enough ability to attribute mental states to oneself and others to understand that others have beliefs, desires, intentions, and feelings that are different from one's own and which lead to correspondingly different courses of action. There is no need to step beyond the allotted suite of cognitive capabilities.

Importantly, much of the plausibility of the lie hypothesis lies in the Mother's reluctance to believe her own moral thoughts. Given that she arrives at moral thoughts in a purely combinatorial way and so regardless of outside input, and that her upbringing would of course have been amoral, it is difficult to see how or why she would come to believe their content and so take them to correctly characterize some real normative state of affairs. Yet, as I will argue in the next section, it is much easier to see why her children would form the corresponding beliefs based on the Mother's testimony.

3.1.6 The onset of lasting moral belief

On the lie hypothesis, the children, like the Mother, possess productive language. They can grasp the meaning of utterly novel complex linguistic expressions on account of their grasp of its structure and the meanings of its constituents. Since the children would also exhibit instrumental rationality and would be similarly versed in modal matters, they would have the requisite grasp of the constituents' meaning. To convey the thoughts, the Mother would just have to express herself using linguistic expressions that are familiar to the children. Given that "the Mother" has as many attempts as copious amounts of time allow, it is plausible to suppose that she would have eventually succeeded. Once communicated, the children come to believe their content, for the Mother communicates the thoughts as fact.

It is plausible to suppose that these beliefs would have been long-lasting, for at least two reasons. First, as far as we can tell, the human societies of the time (~200,000-40,000 years ago) were hunter-gatherer enterprises that relied heavily on apprenticeship to impart foraging skills and knowledge to the young (Sterelny, 2012). Presumably, the Mother would be one of her children's primary informants and so there is likely to have been a pattern of trusting acceptance of her utterances. That in such a context

moral claims concerning proper action would be understood and then promptly accepted is not an implausible suggestion. Second, these were preliterate societies, which – to the best of our knowledge – would only exhibit the capacity for intentional symbolic representation after the start of the Upper Paleolithic Revolution (Bar-Yosef, 2002). There is a great deal of evidence suggesting that our modern metalinguistic awareness of the definitions of words, the logical relations among utterances and even the way utterances sound to us is deeply influenced by our literacy (Olson, 2016). It would be burdensome for individuals in a culture with no knowledge whatsoever of writing or even the possibility of writing to systematically evaluate the truth of claims, let alone moral claims that are not amenable to empirical falsification, let alone as a child. For these reasons, I think it is implausible to suggest that the children would either be too suspicious or too skeptical to believe the Mother.

Furthermore, the Mother's utterances are intended to yield belief that would inform behavior. It is plausible to think that she would eventually succeed, for at least three reasons. First, as I noted in the previous section, her folk-psychological understanding of human behavior is quite sophisticated. She has a basic understanding of belief-desire psychology and enough of an understanding of communication to realize that utterances trigger can trigger belief in their content. She has the knowledge and understanding to construct effective linguistic expressions. Second, the Mother has a great deal of time to arrive at the appropriate expressions, since at the very least there have been tens of thousands of years for the moral beliefs systems of today to develop from these meager beginnings. Lastly, and most importantly, the Mother is to trigger beliefs that are explicitly about actions. If the belief is formed by the children, to behave as though some action is unconditionally wrong or unconditionally right would seem to require, respectively, either always avoiding performing it or performing it whenever pertinent. Since the beliefs are precisely about the unconditional viability of various actions, these will likely to inform decisions among various courses of action.

Now that I have explained in greater detail how the Mother's parental strategy would work and why it is empirically plausible that it would work, I turn to explain how the strategy might have then spread across the population.

3.1.7 The dissemination of the first moral beliefs

On the lie hypothesis, the Mother tells her children that certain actions are unconditionally right or unconditionally wrong because these beliefs modify their patterns of behavior for the better and for the long term. As a behavior modification strategy, it is presumably less effortful, time-consuming, or otherwise costly than, say, corporal punishment or positive reinforcement, since these other parenting techniques require repeated application rather than, possibly, just the one-time formation of belief. The Mother's parental strategy would be a welcomed addition to the human parenting toolkit, even if more frequently than not, initially, parents had to continually remind their children of the moral claim to ensure that it is not forgotten.

Insofar as the Mother's strategy is successful, moreover, it is likely to be adopted by others in her social group. As selection theorists Joseph Henrich and Robert Boyd (2001) have shown, payoff-based transmission (a tendency to copy the most successful individual), together with conformist transmission (a tendency to copy the most frequent behavior), can stabilize cooperative behavior within a group and by another form of selection spread beyond that group to the general population. Since the cognitive mechanisms referred to above can support these types of cultural transmission (ibid., pp. 80-82), this model can explain the spread of the strategy across the population.

Of course, Henrich and Boyd's model might make assumptions that prove to be false, rendering it inadequate as representation of the target evolutionary process. In particular, in relying on their model, I assume that both the Mother's group and subsequent generations possessed a psychological bias towards copying the majority. This may be incorrect, but recall that the lie hypothesis is merely aiming to explain how moral belief *might have* come about, not how it in fact did. Besides, this assumption is empirically supported, since both us and our closest living relatives, chimpanzees, exhibit this bias (Haun, Rekers, & Tomasello, 2012). Since we are so closely related to chimpanzees, it would be unmotivated and ad hoc to supposed that we (the chimps included), evolved the mayority bias independently of one another and in our case *after* the advent of moral belief.

In any case, other mechanisms or models may be capable of explaining the dissemmination of the strategy; it might just be a temporary failure of human imagination that prevents us from modeling the elements of the requisite evolutionary process.

Notice, lastly, that on Henrich and Boyd's model neither the Mother's group nor subsequent generations need to have insight into the usefulness of the strategy, for they just have to first appreciate the fact *that* it is a successful strategy and, later on, merely register its frequency. This is all others' "seeing the brilliance" of the Mother's strategy has to amount to. But how does the lie hypothesis fold into the moral prosociality hypothesis?

3.1.8 From kin morality to human morality

Recall that, according to the moral prosociality hypothesis, moral cognition evolved to play a role in the evolution of human cooperation by promoting sharing and helping behavior and/or deterring cheating and defection among non-kin. The lie hypothesis posits a mechanism for the improvement of cooperation among kin, specifically, parent-offspring cooperation, but it can be easily co-opted to also enable cooperation among non-kin.

Note that the Mother and her imitators had to keep their lies a secret, for otherwise the children's beliefs would not keep. Since at the earliest this occurs 40,000 years ago, long before recorded history, the memory of this deception does not survive more than a few generations, quickly becoming a secret everybody forgot. In time, as the memory fades, and sincere moral belief spreads across the population, some begin to take the very small logical step to the view that an unconditionally right or wrong action is right or wrong, not just for them to perform, but for anyone to perform, for it is wrong no matter what or right come what may. In this way, it is easy to see how – given enough time (and there is plenty) – our ancestors begin to moralize each other *in addition to* their children. If our ancestors have domain-general inferential capabilities, or the cognitive abilities needed for competent deduction,[48] they should

[48] For a "how possibly?" explanation of our reliability of deductive reasoning, see Schechter (2013). Of course, deduction must have been reliable by then for this suggestion to work.

organically come to hold each other accountable for their actions, as they would their children, for this is something they have already done and can repeat.

Don't forget that I am not claiming that this is how the practice of moralizing each other *actually* came about. I aim only to give an account that is compatible with the available empirical evidence and with the background constraints. Right now, I only mean to show that the lie hypothesis is permitted by current evolutionary science. If we can take the cognitive mechanism(s) for competent deduction to be among those present prior to the advent of moral belief, the lie hypothesis feeds into the moral prosociality hypothesis, becoming explanatorily locked in step.

On the lie hypothesis, around the time the cultural evolution of the differentiation of moral beliefs started to kick in – that is, at the dawn of the recruitment of moral cognition for non-kin, cooperation evolution – moral beliefs begin to be selected according to relative prosociality.[49] Since we can assume these mechanisms were present (§1.4), this story of the transition from moralizing children to moralizing each other would explain how an inventive parental strategy might have become fuel for the mechanisms of the moral prosociality hypothesis and thus a staple of human social interaction.

To sum up, then, the lie hypothesis is consistent with the byproduct approach, because it does not presume there to be cognitive abilities that go beyond the allotted suite (§§3.1.4-6), and it is consistent with the moral prosociality hypothesis because there is at least one way the parental strategy can come to be applied to all humans and hence among non-kin (§§3.1.7-8). Since it is also consistent with the available empirical evidence (§3.1.3), though much more would have to be done for a full defense, the lie hypothesis looks to be a genuine candidate "how possibly?" explanation of the advent of moral belief. It's time to turn to the morality as heirloom hypothesis.

3.2 The Heirloom Hypothesis and the First Moral Belief

The hypothesis implicit in the morality as heirloom assumption states that our parents taught us our moral principles, they were taught theirs by their parents and so on up to the advent of moral belief

---

[49] To get a sense of the possible courses of the selection, see Sober (2006).

(call it *the heirloom hypothesis* for short). The label *heirloom* is appropriate because our body of moral knowledge is conceived of as the accumulative, multigenerational, collective achievement of humanity. Accordingly, I assume that in all its versions the heirloom hypothesis allows for the possibility of moral knowledge and takes its actual development to have been cognitively effortful and therefore hard-won. For the sake of argument, moreover, I will assume that this hypothesis is consistent with the available empirical evidence and with the moral prosociality hypothesis and the byproduct approach.

In the following sections, I attempt to fill out the heirloom hypothesis at the advent of moral belief in the hopes of having something to compare to the lie hypothesis. For simplicity, I treat our moral belief systems as unrealistically etiologically pure. Specifically, I will assume that there is a first moral belief that set off the cultural chain reaction that over thousands of generations generated present-day moral belief systems and that, moreover, all subsequent moral belief is ultimately based *solely* on the first. On these assumptions, I can isolate the various possible origin scenarios of our moral belief systems. If I succeed in considering all such possible etiologies of moral beliefs, these simplifying assumptions should not lead us astray, since the aim is to find at least one version of the heirloom hypothesis that can explain the advent of moral belief and if I find none then no combination of origin scenarios should work.

Before proceeding, though, let me explain why the posit of a first moral belief is not ill-defined. As I have been understanding belief formation, it is the real psychological process by which a proposition is believed and thus its meaning represented and treated as if it was true (Gilbert, 1991). If the proposition has a moral subject matter, and it is the first to be believed by a human being, then it is the first moral belief. Since moral belief formation occurs only if these binary conditions obtain, it makes sense to suppose that someone was the first to meet these conditions, even if it may be more realistic to supposed that multiple individuals were involved in the formation of the first few moral beliefs.[50]

---

[50] It is worth noting how cultural evolution differs, in this respect, from biological evolution. There is no first opposable thumb or first eye or first human, but that's because the physical basis of the trait is transmitted across generations and in the process sometimes modified. Unlike the biological evolution of organisms or inherited traits, the cultural evolution of beliefs merely involves the transmission of their representational content, not of beliefs' actual neural basis. In expressing a belief, one individual produces a perceptible utterance that may in turn trigger the

In the following sections, I consider various kinds of naturally occurring symbolic representations that can carry normative or evaluative information and that may have triggered the first moral belief: namely, utterances, occurrent thoughts, and psychological cues.[51] I consider first the possibility that on the heirloom hypothesis the first human to hold a moral belief formed it thanks to the utterances of his or her contemporaries.

3.2.1 The longevity of epistemic luck

On the heirloom hypothesis, could the first moral belief have arisen due to the words of others? The first thing to note about the first moral belief is that *it is the very first*. In consequence, the other's utterance of these words must not express a moral belief. So, either the utterance is (i) insincere, as in the Mother's case, (ii) noncommittal, as in moral conjecture, or (iii) misunderstood or misarticulated, as in speech errors or mishearings.

As the Mother's example makes clear, if the utterance is insincere, then the first moral belief is, if true, luckily true. Think back to the Mother's deception of her children. Since her utterance of p is insincere, even if p is the case, her children's belief that p is not knowledge because there is no non-accidental connection between the fact that p and their evidence for p, their Mother's testimony. It is a lucky accident that the children happen to believe a truth. The children stand no chance of gaining moral knowledge based solely on the Mother's testimony.[52]

What's more, no future generation stands any chance of gaining moral knowledge on this basis, either. Suppose the children grow up and tell *their* children that p. Let's say they do so sincerely. Since

---

corresponding belief in someone else. Unlike in the genetic transmission of traits, in cultural evolution the replication of beliefs occurs in parallel, not serially, and in consequence beliefs are not the bearer of

[51] I will ignore the possibility that moral belief might have originated in dreams. Though certainly a naturally occurring symbolic representation, presumably, moral knowledge cannot be acquired in dreams.

[52] On some views, the process of communicating via testimony does not involve an utterer transmitting his or her belief to an addressee along with the epistemic properties it possesses. Jennifer Lackey, for instance, holds that "a speaker offers a statement to a hearer, along with the epistemic properties *it* possesses, and a hearer forms the corresponding belief on the basis of understanding and accepting the statement in question" (2008, p. 72). Since the statement, if true, would be luckily true, Lackey, too, should view the Mother's utterance to be inadequate for knowledge.

the grandchildren's evidence for *p*, their parents' testimony, is based on the Mother's insincere utterance of *p*, *their* belief that *p* is also not non-accidentally connected to the fact that *p*, for their parents' testimony is no less luckily connected to that fact than that of the Mother. However many generations go on to pass, if all moral belief is ultimately based on the Mother's testimony, none will ever amount to knowledge, for no subsequent moral belief will be less lucky to be true than the first. Since the heirloom hypothesis assumes the possibility of moral knowledge, the first moral belief cannot be due to an insincerity.

Now, suppose the Mother's articulation of *p* is qualified to imply that she does not believe that *p*. If she signs or says, "I wonder whether *p*" (or some other utterance that articulates *p* but doesn't indicate belief in *p*) and on that basis her children come to believe that *p*, then their belief that *p* is, if true, luckily true and so does not amount to knowledge. For the same reason as above, in subsequent generations it will never be the basis of knowledge. On the heirloom hypothesis, moral belief cannot originate in a noncommittal source. Furthermore, if others' input is scrambled due to errors in either utterance comprehension or articulation, then the first moral belief, if lucky enough to be true, would even more obviously be epistemically luckily true. For the same reason as above, the truth of no subsequent moral belief will be any less lucky than that of the first, if ultimately formed on its basis. Again, moral knowledge cannot originate in a failure of communication, either.

Interestingly, many prominent philosophers hold, or find to be quite plausible, the view that moral knowledge can be attained even if our culture luckily converges on roughly true moral beliefs.[53] This view usually goes undefended when mentioned, on account of its supposed intuitive plausibility. Here, for example, is Roger White:

> Why can't we recognize evaluative truth when we stumble upon it even by extraordinary accident? That we have an obligation to care for our children for instance seems about as easy to recognize as anything. (2010, p. 589)

Gideon Rosen, for another, writes,

---

[53] An exception is Kieran Setiya. See his *Knowing Right from Wrong* for a dissenting view.

> I see no reason this procedure [according to which we learn moral principles from our parents which we then refine and revise in light of experience and reflection according to a critical discipline which we also inherit] should not yield knowledge of transcendent moral principles, even though our reliability in the area would then be contingent on the lucky convergence of our culture on roughly true moral precepts. (1998, p. 398)

Rather, it seems to me that there is no reason to suppose that it *could* yield knowledge, since even if we do attain this kind of accidental reliability our many true moral beliefs would still be luckily true and so no better suited to amount to knowledge. Notice, moreover, that, however convincing or firm the semblance of recognition is in subsequent generations, if the original belief in a moral obligation to care for one's children were due to an insincere or noncommittal utterance or to a failure of communication, it would not be knowledge and so – even if accurate – not recognition at all. Further, it should be obvious that the number of luckily true beliefs does not change how luckily true they are or why their luckiness prevents them from amounting to knowledge. Such a "more, the merrier" epistemic principle is plainly absurd. But if that's not what is behind these comments, then what is?

Rosen's comments point in a few directions. Perhaps the hidden bit of reasoning is that because moral cognition would tend to yield true moral beliefs it would therefore be able to generate knowledge. But think back to the lie hypothesis. If true, it continues to be possible albeit quite unlikely that, nevertheless, over many generations, as a matter of happenstance, human moral cognition comes to reliably output true moral beliefs, our culture luckily converging on roughly true moral principles. Yet, intuitively, this is no improvement on our epistemic predicament with respect to moral matters. For the Mother's testimony, the original basis of all moral beliefs, is no indication of their truth and so our believing in many moral truths is no less an accident and hence no closer to knowledge. The fact that our culture is such that more of our guesses are true does not make mere guesswork amount to knowledge, however dogmatic or unwitting the guesswork.

Perhaps it is the role of the critical discipline that we inherit alongside our moral beliefs that makes the difference. Maybe Rosen thinks that all the epistemic action lies in the refinement and revision of our moral principles in light of experience and reflection. But whether any revision or refinement

improves our moral epistemic situation still hinges on whether we have some starting body of knowledge to develop in accordance with the tenets of this discipline. If human beings have no moral knowledge to begin with (as our ancestors, lacking in all moral belief, wouldn't), it is far from obvious how revision or refinement, carried out by those without a clue, would ever gain the moral knowledge to impart upon us. Neither blind revision nor blind refinement can make a fool knowledgeable, no matter how effectively it may inflate the fool's presumption of knowledge.

If the first moral belief came about thanks to the words of others, the heirloom hypothesis cannot explain the advent of moral belief, for on this possibility there would be no knowledge to pass on across the generations. Furthermore, there should not be a presumption in favor of the possibility of moral knowledge in case of a lucky convergence on the moral truth. Nevertheless, epistemic luck may in other ways be avoided, for as Rosen notes we do in fact revise our moral beliefs in light of reflection and experience. Perhaps the first human to hold moral belief arrived at it based on his or her experience or reflection. I turn to that possibility next.

3.2.2 The nature and necessity of evolutionary novelty

On the heirloom hypothesis, could the first moral belief arise on its own? The second thing to note about the first moral belief is that it is the first *of its kind*. Not only had it never been held, *no one else had formed beliefs with such a subject matter*.

To plausibly claim that the heirloom hypothesis hits its explanatory target, it must include some explanation of how or why the posited first moral belief is to be regarded as concerning an evolutionary novel subject matter. This desideratum is different from the above methodological requirement that the posited first moral belief must be plausibly hypothesized to transition into the moral beliefs we hold today. This much I grant in assuming for argument's sake that the heirloom hypothesis is consistent with background constraints. Rather, the desideratum is that, as a hypothesis of the advent of *moral* belief, intuitively in a class of its own, some factor must be posited to explain why the first moral belief is about a brand new subject matter rather than an innovation in thinking on an older matter. This feature should not only desired in, but expected of, any account of the development of evolutionarily novel forms of cognition, human or otherwise.

To get a better sense of this explanatory demand, I'll explain how the lie hypothesis meets it. Recall that on the lie hypothesis the Mother produces moral thoughts combinatorially, specifically, by jointly exercising evolutionarily novel combinations of cognitive abilities. So, equipped with productive language and hence productive cognition, the Mother combines pre-existing ideas in creative enough ways to produce thoughts with moral subject matter. Equipped with the species-typical endowment, her children can form beliefs with these thoughts' content. Since these propositions are communicated to the children as fact, trusting of their Mother, they come to hold the first moral beliefs.

On the lie hypothesis, it is the joint, coordinated exercise of cognitive abilities that *have never been exercised in this way before* which yields belief with an evolutionarily novel subject matter. To see why this might explain the novelty of the subject matter as opposed to merely the novelty of the belief, consider what is perhaps the most famous example of a thought formed with an eye to novelty:

Colorless green ideas sleep furiously.

This is Chomsky's (1957) example of a grammatically correct but semantically nonsensical sentence. For present purposes, I take it to express a well-formed thought, if one empty of any clearly understandable meaning. If in productive cognition the combination of old ideas can yield well-formed but nonsensical thoughts, then surely well-formed thoughts of novel subject matter are not beyond its bandwidth. After all, productive cognition is standardly supposed to be discretely infinite, that is, to "make infinite use of finite means," in 18th century Prussian philosopher and linguist Wilhelm von Humboldt's apt phrase.[54] In any case, productive cognition certainly *might have* been implicated in the advent of moral belief.

I hope the need to explain the novelty of the first moral belief's subject matter is by now clear. If on the heirloom hypothesis the first human to hold a moral belief forms it on his or her own, then whatever the precise account it must involve the evolutionarily novel, joint, coordinated exercise of the human cognitive endowment. Since its exercise may be either voluntary or involuntary, in the next two sections I consider whether these possibilities represent an opportunity to fill out the heirloom hypothesis

---

[54] But see Pullum and Scholz (2010) for a critique of this infinitude claim.

at the advent of moral belief, beginning with the possibility that the first moral belief ultimately came about voluntarily.

3.2.3 The combinatorial structure of productive cognition

I can only think of one way the first moral belief may have ultimately resulted from the voluntary use of cognitive abilities: the first human to hold a moral belief intentionally combined ideas to form moral thoughts and later came to believe in their content. Since productive cognition is the only known capacity to yield thoughts on novel subject matters (Corballis, 1991; Boden, 2004), I doubt this will be too significant a failure of my imagination. So, could on the heirloom hypothesis someone creatively combine ideas to yield moral thoughts and later believe their content?

Suppose the first moral belief could have arisen in this way. Since the capacity for productive cognition quite probably operates combinatorially (Fodor & Pylyshyn, 1988), the first human to hold moral belief (for short call him or her *the maverick*) must have combined ideas to produce moral thoughts in one of three ways: either (i) in indifference to the captured proposition, as in Chomsky's case, (ii) in indifference to the captured proposition's truth, as in the Mother's case, or (iii) with the aim of capturing a truth, as in communication via testimony under oath.

First consider Chomsky's case. In arriving at the thought that colorless green ideas sleep furiously he was aiming to distinguish between the grammaticality and the meaningfulness of a sentence. Any thought that helps draw this distinction, whatever its propositional content, serves this purpose. It did not have to be about colorless green ideas or furious sleep. The production of this thought proceeds, in this respect, in indifference to the proposition to be represented by the resulting thought, since there was nothing it had to be about. If the maverick forms moral thoughts in this way, it is clearly a matter of luck if any turn out to represent a truth, since the implicated combinatorial process would have its propositional content by accident, to say nothing of its truth. However the maverick goes on to believe these thoughts' content, if his or her only experience with moral matters are these thoughts, then such a belief would not amount to knowledge nor, for reasons given above, could it be the basis of knowledge in subsequent generations. Moral knowledge cannot originate in such a content-indifferent source.

Now, consider the Mother. In arriving at the thought that a certain action is unconditionally right or wrong, she is aiming to tell a believable and relevantly behavior-guiding lie. In consequence, the resulting thought must express a proposition about acts and their unconditional rightness or wrongness. In this sense, its production does not proceed in indifference to the target proposition, for there is something it *must* be about. But since the aim is merely to represent a believable (as opposed to plausible) proposition, belief in which would merely have the desired behavioral profile, the production of the thought proceeds, in this respect, in indifference to the target proposition's truth. Hence the Mother's testimony's epistemic fruitlessness: it cannot support a non-accidental connection to the fact it would represent, even if truth-conducive, because it stumbles upon the truth accidentally. Nothing of epistemic relevance changes if instead of merely telling her lies the Mother believes them. If the first moral belief is formed in this way, like the children's, it can neither amount to knowledge nor be its basis in future generations. Moral knowledge cannot originate in this kind of a truth-indifferent source.

Lastly, suppose that on the heirloom hypothesis the maverick attempts to produce his or her first moral thought with the aim of capturing a truth. This is the most promising of the three possibilities. Unlike the other two, that belief in the thought's content would develop can much more comfortably be taken for granted. What's more, for argument's sake, I'll make the questionable assumption that such a radically novel thought could have been produced with such an aim in mind. Nevertheless, since this would be his or her first thought with this subject matter, there would be no background constraints on the combinatorial process responsible for it and so none of the true beliefs needed for the reliability of this process would be present.

Take the topic of the advent of moral belief. Suppose that my first thought about it was that the advent of moral belief must have occurred naturally. In arriving at this thought, I leaned on the theoretical framework of evolutionary science, most crucially, on the tenets of common descent and natural (as opposed to guided) evolution. Had I not held these views, this thought would have been luckily true. Since I would not have been privy to any of the relevant background information, whether I had been aiming to think a truth or not, I would not have succeeded in doing so non-accidentally. Likewise, the maverick might aim to form a true thought, but since there would be no pre-existing moral belief, the background to pursue this aim would not be in place, for there would be no conceptual framework to

help specify the thought's content in a truth-conducive manner. So, however heartfelt the attempt to capture a truth, both thought and corresponding belief would fail to be a basis for knowledge and for the familiar reasons fail forever.

It may be objected that since other sorts of normative beliefs probably predated moral belief it would be too quick to conclude that nothing could have helped guide the production of the first moral thought in a truth-conducive manner. Perhaps, the conceptual paraphernalia for reliable moral cognition was present in an inchoate form, ready to be cobbled together in subsequent generations. I freely admit that this is a possibility that should be taken seriously. But until there is a concrete suggestion as to how these pre-existing normative concepts might have come together, where they might have come from and why their combination might lead to reliable moral belief formation, there is no reason to think that such a thing *might have* occurred. Indeed, under the present description, for all we know, this development in human cognition has yet to pass, even if possible. So, while I agree there is this formal possibility, undeveloped it does nothing to improve the explanatory standing of the heirloom hypothesis.

The heirloom hypothesis cannot explain the advent of moral belief on the assumption that the first moral belief ultimately resulted from the voluntary use of cognitive abilities. Last I turn to consider the possibility that on the heirloom hypothesis the first human to hold moral belief arrived at it on his or her own and in an entirely involuntary manner.

3.2.4 Psychological cues and moral belief formation

If the first moral belief was not formed in response to others' utterances or one's combinatorially formed thoughts, it must have been due to causal interaction with the environment, more specifically, the result of perceiving a *psychological cue*, that is, a hint or guiding suggestion as to how to act, where an *environmental cue* is an event and a *behavioral cue* is an action. While, obviously, actions are events, the difference here is in how each type of cue is perceived: either as issuing from someone's agency or as not. A raise of the brow may signal the wisdom of proceeding with caution and a drowning child the need for heroic action.

Having moved beyond utterances and thoughts, I know of no other naturally occurring symbolic representation but cues, that can convey normative or evaluative information. For argument's sake, I'll assume that the perception of cues is sufficient for the uptake of their content. Since inference's involvement would implicate pre-existing beliefs on inferentially related subject matters, to assume otherwise would make the suggestion of cues a non-starter. So, could on the heirloom hypothesis the first moral belief result from either a behavioral cue or an environmental cue?

Suppose the first moral belief is formed on perceiving a behavioral cue and that you are the one to form this belief and I am this cue's agent. Since I don't have any moral beliefs, either you misinterpret the significance of the cue or I misled into thinking it conveys this information. Since prior to forming your first moral belief you have no credible idea of morality, and neither do I, whether you misinterpret or I misled, your belief would not be non-accidentally connected to the fact represented, thereby failing to become knowledge and for reasons familiar from above incapable of grounding future knowledge. So, even on the hefty assumption that a behavioral cue, a non-language-like representation, can convey information about an evolutionarily novel subject matter,[55] it cannot yield moral knowledge and so this kind of an origin scenario cannot be accommodated by the heirloom hypothesis.

Now, suppose the first moral belief is formed upon perceiving an environmental cue. Hard as it is to imagine what this cue could be, we know a few things about what it must have been like based on the discussion so far. Since the first moral belief is about an evolutionarily novel subject matter, its formation requires the evolutionarily unprecedented co-activation of the enabling cognitive mechanisms. In consequence, this cue must be such that its perception results in the complex mix of cognitive activity proper to moral belief formation. Since this mix of activity would have only then been achieved for the first time, no one would have perceived such a cue before then, for the belief is formed involuntarily and they would have had the same species-typical cognitive endowment. Therefore, the first human to hold

---

[55] This is a gigantic assumption. Though cue-based systems of human communication can be compositionally structured, as in the facial expression of emotion (Du, Tao, & Martinez, 2014), none are discretely infinite and so communicating ideas about evolutionarily novel subject matters is very probably beyond its capabilities.

moral belief was either the accidental beneficiary of just the right confluence of environmental factors or the first to register an evolutionarily novel aspect of the environment.

On the first scenario, the first moral belief is an accidental byproduct of the individual's presence in an evolutionarily unprecedented and unlikely environment and, on the second, it is the product of the individual's keen observation of his or her surroundings. On neither origin scenario must the first moral belief, if true, be luckily true. While on the first there could have easily been no moral beliefs (since the improbable cue could have easily failed to materialize), if this belief is formed based on nonmoral facts that are relevant to its truth, then the connection between the basis for belief and its truth would be non-accidental. There would be luck in that moral belief happened to be formed based on the morally relevant nonmoral facts (or, indeed, that they were formed at all), but it would be the kind of "luck of the draw" involved in accidentally gathering good, serviceable evidence. (Cf. Pritchard 2007)

On the second scenario, the environmental cue is a relatively recent product of evolution, cultural or biological, that lead to a change in either or both the human way of life or the environment. The first human to hold moral belief was the first to note the change and in consequence became the first to hold moral belief. If this belief was formed based on the evidentially relevant nonmoral facts, then the connection between the basis for this belief and its truth would be non-accidental. Again, there would only be luck in that moral belief turned out to be formed on right sort of truth-conducive basis and hence involves only the luck of the draw, not epistemic luck. So, on this origin scenario, moral knowledge would be possible. Since the aim was just to supply the heirloom hypothesis with an account of how moral belief *might have* originated, this and the previous origin scenario suffice for present purposes. To explain the advent of moral belief, the heirloom hypothesis must posit an origin scenario that includes the perception of environmental cues. It's time to compare the lie hypothesis to the heirloom hypothesis.

3.3 The Heirloom Hypothesis vs. the Lie Hypothesis

I should make it clear at the outset that I'd be lying if I told you that morality is a lie—or for that matter an heirloom. That is plainly beyond the available evidence. The question is which hypothesis presents the more empirically plausible proposal. If a stand must be taken, should we take morality to be a precious heirloom or a tremendous lie? In answering this question, I'll continue to grant for argument's

sake that the heirloom hypothesis is consistent with the empirical evidence and with the background constraints and that the perception of environmental cues suffices for the uptake of their content. So, which is it? Lie or heirloom?

3.3.1 Evolutionary novelty, revisited

On the lie hypothesis, the first moral belief is about an evolutionarily novel subject matter because the Mother creatively combines pre-existing ideas to generate thoughts with novel subject matters, utilizing her capacity for productive cognition. On the heirloom hypothesis, by contrast, since the first moral belief forms involuntarily, the first human to hold moral belief does not arrive at it creatively, since he or she has no control, direct or otherwise, over the cognitive process that produces this belief. While the lie hypothesis can appeal to the discrete infinity of productive cognition to explain the novelty of the first moral belief, the heirloom hypothesis is forced to stipulate that the environmental cue is such that it can explain this novelty, having no recourse to discrete infinity. In this respect, the lie hypothesis is explanatorily superior to the heirloom hypothesis.

3.3.2 The genesis of moral content

The lie hypothesis is also explanatorily superior in another respect. Since the Mother intentionally generates moral thoughts with a specific aim in mind, there is a specific story about how her children came to form moral beliefs with precisely that content. But on the heirloom hypothesis, since it is the perception of some underspecified environmental cue (or perhaps sequence of cues) that triggers the first moral belief, there is no sense of why it ended up having the content it did. The lie hypothesis provides the fuller "how possibly?" explanation of the formation of the first moral beliefs.

3.3.3 Metaphysical impartiality

The lie hypothesis has a further advantage: it can explain the advent of moral belief whatever the correct metaphysical view of morality. Since both moral thought production (§2.3) and moral belief transmission (§§1.5 & 1.6) may proceed without regard to moral truth, the lie hypothesis can explain how moral belief might have first been formed whatever you think of its nature or existence. But, as we

saw throughout §2, on the heirloom hypothesis, because of its commitment to the possibility of moral knowledge, the background metaphysical view can interfere with its explanatory potential. Indeed, the heirloom hypothesis is inconsistent with a mind-dependent view of moral facts, since moral knowledge is thought of as hard-won, as the cumulative, multigenerational achievement of humankind.

3.3.4 The possibility of moral knowledge

The lie hypothesis is, however, inconsistent with most conceptions of moral knowledge, whereas the heirloom hypothesis takes the possibility of moral knowledge as a given. On the one hand, as we saw in §2, this assumption can be a liability, for the heirloom hypothesis could not accommodate most origin scenarios and in consequence had to rely on brute stipulation to ensure that it hit its explanatory target. On the other hand, we routinely attribute moral knowledge to ourselves and each other, and not of the subjective kind involved in learning about mind-dependent moral facts. Prior to evolutionary theorizing, there should be, it seems, a presumption in favor of the possibility of objective moral knowledge. But as we saw in §2.1, this presumption does not survive evolutionarily informed epistemological scrutiny.

3.3.5 And the winner is ...

I think that we can only conclude that the lie hypothesis is the superior proposal. Not only can we count the above explanatory advantages in its favor, we can discount one apparent disadvantage, the implied impossibility of objective moral knowledge. Furthermore, let's not forget that I have only been assuming for argument's sake that the heirloom hypothesis is consistent with the background constraints and that the perception of cues is sufficient for their uptake. These assumptions must be substantiated before the heirloom hypothesis can be a genuine candidate explanation, even in outline. I think that to the best of our knowledge and understanding, human morality is a lie rather than an heirloom.

3.4 Conclusion

First I outlined and defended the lie hypothesis, explaining along the way why it is consistent with the empirical evidence available, how it relies on the byproduct approach, and why and how it easily folds into the moral prosociality hypothesis. The Mother lied to get her children to behave, her success

gained her imitators, her imitators' children never learned her secret, her secret forgotten her tactics came to be applied among peers. I then tried to fill out the heirloom hypothesis at the advent of moral belief: first by appeal to others' utterances, then to one's own creativity, and finally to the cues all around. Only the environmental cues could potentially support a non-accidental connection to the moral facts, severely weakening its appeal, giving the heirloom hypothesis instead an ad hoc feel. To the best of our knowledge, human morality is not an heirloom but a lie.

In the end, though, I hope to have convinced you of just one thing: that to *know* whether human morality is a phantom of misbelief or a hoard of epistemic treasures, we must venture beyond the armchair to inquire its evolutionary origins.

References

Alexander, R. D. (1987). *The biology of moral systems.* Chicago: Aldine Transaction.

Ayala, F. J. (2010). The difference of being human: Morality. *Proceedings of the National Academy of Sciences*, 9015-9022.

Bar-Yosef, O. (2002). The archeological framework of the Upper Paleolithic Revolution. *Annual Review of Anthropology*, 363-393.

Boden, M. A. (2004). *The creative mind: Myths and mechanisms* (2nd ed.). London: Routledge.

Boyer, P. (2002). *Religion explained: The evolutionary origins of religious thought.* New York: Basic Books.

Boyer, P. (2003). Religious thought and behaviour as by-products of brain function. *Trends in Cognitive Science*, 119-124.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Chomsky, N. (1969). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Corballis, M. C. (1991). *The lopsided ape: Evolution of the generative mind.* Oxford: Oxford University Press.

D'Anastasio, R., Wroe, S., Tuniz, C., Mancini, L., Cesana, D. T., Dreossi, D., . . . Capasso, L. (2013). Micro-biomechanics of the Kebara 2 hyoid and its implications for speech in Neanderthals. *PLoS ONE.*

Darwin, C. (1879). *The descent of man.* London: Penguin Books.

de Waal, F. (2009). *Primates and philosophers: How morality evolved.* Princeton: Princeton University Press.

de Waal, F., Churchland, P., Pievani, T., & Parmigiani, S. (Eds.). (2014). Evolved morality: The biology and philosophy of human conscience [Special issue]. *Behaviour, 151*(2-3).

Deem, M. J. (2016). Dehorning the Darwinian dilemma for normative realism. *Biology and Philosophy*, 727-746.

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, E1454-E1462.

Dunbar, R. I. (1998). The social brain hypothesis. *Evolutionary anthropology*, 178-190.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wiebe, V., Kitano, T., . . . Paabo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, 869-872.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 3-71.

Gibson, K. R., & Tallerman, M. (Eds.). (2011). *The Oxford handbook of language evolution.* Oxford: Oxford University Press.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, 107-119.

Godfrey-Smith, P. (1996). *Complexity and the function of mind in nature.* Cambridge: Cambridge University Press.

Gould, S. J. (1991). Exaptation: A crucial tool for an evolutionary psychology. *Journal of Social Issues*, 43-65.

Gould, S. J., & Vrba, E. S. (1982). Exaptation: A missing term in the science of form. *Paleobiology*, 4-15.

Haun, D. B., Rekers, Y., & Tomasello, M. (2012). Majority-biased transmission in chimpanzees and human children, but not oranguntans. *Current Biology*, 727-731.

Hauser, M. D. (2006). *Moral minds: How nature designed our universal sense of right and wrong.* New York: Ecco.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 1569-1570.

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., . . . Lewontin, R. C. (2014). The mystery of language evolution. *Frontiers in Psychology*, 1-12.

Henrich, J., & Boyd, R. (2001). Why people punish defectors. *Journal of Theoretical Biology*, 79-89.

Houston, S. D. (2004). *The first writing: Script invention as history and process.* Cambridge: Cambridge University Press.

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

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

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

Joyce, R. (2014). The origins of moral judgment. *Behaviour*, 261-278.

Kitcher, P. (2006). Evolution and ethics: How to get here from there. In F. de Waal, *Primates and philosophers: How morality evolved* (pp. 120-139). Princeton: Princeton University Press.

Kitcher, P. (2011). *The ethical project.* Cambridge, MA: Harvard University Press.

Kitcher, P. (2012). Precis of The Ethical Project. *Analyse &Kritik*, 1-19.

Lackey, J. (2008). *Learning from words: Testimony as a source of knowledge.* Oxford: Oxford University Press.

Mandelbaum, E. (2014). Thinking is believing. *Inquiry*, 55-96.

Pritchard, D. (2007). *Epistemic luck.* Oxford: Oxford University Press.

Pullum, G. K., & Scholz, B. C. (2010). Recursion and the infinitude claim. *Recursion in human language*, 113-138.

Rosen, G. (1998). Blackburn's Essays in Quasi-Realism (New York: Oxford University Press). *Nous*, 386-405.

Schechter, J. (2013). Could evolution explain our reliability about logic? (T. S. Gendler, & J. Hawthorne, Eds.) *Oxford Studies in Epistemology, IV*, 214-239.

Setiya, K. (2012). *Knowing right from wrong.* Oxford: Oxford University Press.

Sober, E. (2006). Models of cultural evolution. In E. Sober (Ed.), *Conceptual issues in evolutionary biology* (pp. 535-551). Cambridge, MA: MIT Press.

Sober, E., & Wilson, D. S. (1999). *Unto others: The evolution and psychology of unselfish behavior.* Cambridge, MA: Harvard University Press.

Sterelny, K. (2003). *Thought in a hostile world: The evolution of human cognition.* Malden, MA: Blackwell Publishing.

Sterelny, K. (2012). *The evolved apprentice: How evolution made humans unique.* Cambridge, MA: MIT Press.

Tomasello, M. (2016). *A natural history of human morality.* Cambridge, MA: Harvard University Press.

Trinkaus, E. (2005). Early modern humans. *Annual Review of Anthropology*, 207-230.

White, R. (2010). "You just believe that because ...". *Philosophical Perspectives*, 573-615.