Andreas Mogensen

All Souls College, Oxford

Faculty of Philosophy

Evolutionary debunking arguments in ethics

A thesis submitted for the degree of D.Phil in Philosophy

Hilary Term 2014

University of Oxford

Andreas Mogensen,

All Souls College, Oxford; Faculty of Philosophy

Evolutionary debunking arguments in ethics

D.Phil Submission, Hilary 2014

Abstract:

I consider whether evolutionary explanations can debunk our moral beliefs. Most contemporary discussion in this area is centred on the question of whether debunking implications follow from our ability to explain elements of human morality in terms of natural selection, given that there has been no selection for true moral beliefs. By considering the most prominent arguments in the literature today, I offer reasons to think that debunking arguments of this kind fail. However, I argue that a successful evolutionary debunking argument can be constructed by appeal to the suggestion that our moral outlook reflects arbitrary contingencies of our phylogeny, much as the horizontal orientation of the whale's tail reflects its descent from terrestrial quadrupeds.

An introductory chapter unpacks the question of whether evolutionary explanations can debunk our moral beliefs, offers a brief historical guide to the philosophical discussion surrounding it, and explains what I mean to contribute to this discussion. Thereafter follow six chapters and a conclusion. The six chapters are divided into three pairs. The first two chapters consider what contemporary scientific evidence can tell us about the evolutionary origins of morality and, in particular, to what extent the evidence speaks in favour of the claims on which debunking arguments rely. The next two chapters offer a critique of popular debunking arguments that are centred on the irrelevance of moral facts in natural selection explanations. The final chapters develop a novel argument for the claim that evolutionary explanations can undermine our moral beliefs insofar as they show that our moral outlook reflects arbitrary contingencies of our phylogeny. A conclusion summarizes my argument and sets out the key questions that arise in its wake.

Word count: 74,900

Acknowledgments:

For helpful comments and discussion of my work in this essay, I am grateful first and foremost to my supervisors, Krister Bykvist and John Hawthorne, and to Matthew Braddock, Daniel Deasy, Hilary Greaves, Cecilia Heyes, William MacAskill, Walter Sinnott-Armstrong, Amia Srinivasan, and Ralph Wedgwood. I thank All Souls College for supporting my research. Last but not least, I am indebted to Nicola Mastroddi for her love and support.

Evolutionary debunking arguments in ethics

Table of contents

0.

Introduction

1. Introduction

Can evolutionary explanations debunk our moral beliefs? In this essay, I offer new reasons to think that this question should be answered in the affirmative and that previous arguments to that effect have been flawed. In this introduction, I'm going to unpack the question, offer a brief historical guide to the philosophical discussion surrounding it, and explain what I mean to contribute to this discussion.

The unpacking will be done in section 2, beginning with an outline of some relevant epistemological presuppositions, moving on to consider what qualifies as an evolutionary explanation, and finally addressing whether we should understand our question as encompassing all moral beliefs or just a particular subset. In section 3, I then discuss the history of evolutionary debunking arguments from Darwin to today, noting an important shift in focus with respect to the question of why evolutionary explanations are considered debunking. Finally, in section 4, I outline the aims and structure of this essay.

2. Unpacking the question

2.1 Epistemological presuppositions

We are to consider whether our moral beliefs can be *debunked* by evolutionary explanations. What is it for a belief to be debunked in this sense? My emphasis throughout this essay will be on how evolutionary explanations reflect on the *justification* (or lack thereof) of our moral beliefs. I assume that the question of

justification is intimately connected to normative and deontic questions about what we *ought* to believe and what we are *permitted* to believe. In particular, I assume that being justified in believing *p* is a necessary condition for being permitted to believe *p* and for it to be the case that you ought to believe *p*.

In choosing to focus on justification, I am to some extent attempting to impose order. Epistemic evaluation is not confined to questions of justification: we can ask whether we *know* the things we believe, whether our beliefs reflect various *epistemic virtues* and *vices*, and so on. Whereas some philosophers who believe that evolutionary explanations impact the epistemic status of our moral beliefs are clear that they're concerned with justification, others are not so easy to pin down;[1] some of their critics have chosen to frame the issue in terms of knowledge, rather than justification.[2] I believe that where there is uncertainty, a reasonable interpretation of pro-debunking arguments nonetheless finds the issue of justification in play. I can only really substantiate this exegesis in the ensuing chapters, through my discussion of the various extant debunking arguments. However, I can note the following already here. The debunking power of evolutionary explanations attracts philosophical attention because of the felt possibility that we may be forced to make substantial revisions in our moral outlook in light of new discoveries about the evolutionary origins of morality. For that to be so, it must be the case that evolutionary explanations can speak to what we ought and ought not to believe. As I understand the notion of epistemic justification, they can do so only by addressing the justificatory status of our moral beliefs.

I'm going to further interpret the notion of debunking via the well-known idiom of *epistemic defeat*, introduced by John Pollock (1970, 1986). Thus, evolutionary explanations are debunking iff they supply defeaters. A defeater, roughly speaking, is a

[1] Joyce (2001, 2006) is admirably clear; Ruse (1986) is not.
[2] See, *e.g.*, Brosnan (2011), Wielenberg (2010).

condition that provides a *prima facie* reason to withhold belief from some proposition that we would otherwise have been justified in believing.[3] Defeaters are *prima facie* reasons to withhold belief: they can themselves be defeated by way of conditions known as *defeater-defeaters*. (There are also *defeater-defeater-defeaters*, and so on.)

Two classes of defeaters are widely recognized: *rebutting* and *undercutting*. A rebutting defeater is a reason to believe the contrary of what you believe (or have justification for believing); an undercutting defeater works instead by casting doubt on the trustworthiness of the grounds for your belief (or the method by which it was formed). Thus, the discovery that evidence which we took to support some hypothesis has been doctored will undermine belief in the hypothesis without necessarily providing any positive reason to think the hypothesis is false: this constitutes a merely undercutting defeater. Some defeaters are *hybrid*: they both rebut and undercut. It's plausible that many rebutting defeaters are in fact undercutting as well: if I receive evidence that some method of belief-formation I've used gives the wrong result (a rebutting defeater) and I have no reason to expect that the method delivers the wrong result due to random error or noise, I receive evidence that the method is systematically biased toward error (an undercutting defeater).

Interpreted in the idiom of defeat, the question we are considering asks, in effect, whether evidence we receive about the evolutionary origins of our moral beliefs could affect their justificatory status so as to cancel any pre-existing entitlement that we might have for holding those beliefs. Thus, I'll assume that we are *prima facie* entitled to rely on the relevant beliefs, as well as the methods by which they are formed, including basing our beliefs on our moral intuitions.[4] Obviously, if the beliefs that admit of evolutionary explanations are unjustified on independent grounds, there

[3] A more exact definition of epistemic defeat requires resolving various complications that needn't concern us here. See the excellent discussion in Kotzen (ms.).

[4] For a defence of intuitions as conferring justification on corresponding beliefs see Huemer (2005). Unlike Huemer, I do not presuppose that intuitions are intellectual seemings, as opposed to, say, inclinations to believe. For further discussion of the role of intuitions in moral psychology see ch. 2.

is little point in worrying about the debunking power of evolutionary considerations. We should get rid of them in any case.

Another question that we might have considered, but which I set aside, is that of whether evolutionary explanations could simply *show* our beliefs to be and have been unjustified: whether certain facts about the evolution of human moral psychology might make it the case that our moral beliefs are unjustified, quite apart from the capacity of our awareness of these facts to defeat any prior *prima facie* entitlement. Philosophers who accept some form of *Epistemic Internalism* would naturally reject this possibility: they take it that the justificatory status of our beliefs supervenes on factors 'internal' to the epistemic agent, which rules out any relevance for facts about the distal causes of our beliefs (apart from our awareness of them). Philosophers who accept *Epistemic Externalism*, however, would not reject this possibility out of hand. [5]

There are only so many issues that can be addressed in an essay of this kind. Therefore, my hope has been to distil the central question into a form that's maximally inclusive, allowing me to speak to (almost) everyone at once. Framing the issue in terms of defeat achieves this end, I believe. Both internalists and externalists can get on board with the notion of defeat,[6, 7] and certain kinds of defeaters ought to be relatively uncontroversial between them. For example, evidence that one's belief is formed via an unreliable method is widely regarded as defeating, even by those who do not regard *de facto* reliability as necessary for justification.[8] There might well be cases in which a would-be defeater of this kind is counted by the externalist as evidence that

[5] For example, according to Bergmann's *Proper Functionalism* (Bergmann 2006) a belief is justified only if it is the output of a cognitive mechanism whose function is to produce true beliefs. As Bergmann notes, one prominent account of function is the *Selected-Effect Account*, according to which, roughly speaking, the function of a trait is to produce whatever effect has led to selection for that trait within the organism's evolutionary history (Millikan 1984b, 1989; Neander 1991a, 1991b; Wright 1973; Cf. Buller 1998). Given *Proper Functionalism* and the *Selected-Effect Account*, one might try to argue that our moral beliefs are not justified simply because human moral psychology has not been shaped by selection for accuracy.

[6] See Bergmann (1997).

[7] There may be exceptions: see Hawthorne & Srinivasan (2013).

[8] See Pryor (2000).

your belief was never justified in the first place: evidence of unreliability is a case in point. However, the evidence could be misleading: it might be that your method is reliable after all. Even in that case, the evidence ought plausibly to be classified by the externalist as defeating: evidence of unreliability is defeating, even if misleading.[9] Thus, framing the question in terms of defeat allows both internalists and externalists to participate in the discussion. By phrasing the issue in the way I've done, I hope thereby to have created a unified topic that's as open as can be to people of different epistemological persuasions.

Finally, let me note that our question assumes that our moral commitments are correctly described as *beliefs* and can be assessed, like other beliefs, according to epistemic criteria: in particular, as *justified* and *unjustified*. These assumptions belong, I think, to a common-sense view of morality. Some philosophers in the *Non-Cognitivist* tradition have rejected these assumptions, but they belong increasingly to the intellectual past. More recent work in the *Non-Cognitivist* tradition has aimed to reclaim these assumptions. Proponents of *Quasi-Realism* such as Simon Blackburn (1993, 1998) and Allan Gibbard (2003) suppose there is a 'minimal' sense of belief relative to which our moral commitments *are* beliefs.[10, 11] They have also sought to vindicate a view on which moral judgments can be assessed according to familiar epistemic criteria.[12]

2.2 Evolutionary explanations

The question we're interested in concerns how we should react to evidence that our

[9] Cf. Goldman (1979).
[10] See Blackburn (2010), Gibbard (2003: 180-184). See Sinclair (2006) for valuable discussion.
[11] Cf. Horgan & Timmons (2006).
[12] See Blackburn (1996), Gibbard (2003: 199-287).

moral beliefs can be explained in evolutionary terms. It is not otherwise concerned with how evolutionary biology might impact on our moral commitments.

What exactly is an evolutionary explanation? For many, what most readily comes to mind is likely to be a particular kind of explanation in terms of natural selection: we explain the current prevalence and/or existence of some trait by pointing to its beneficial effects on ancestral fitness. Thus, we might explain why black-headed gulls quickly dispose of empty eggshells by pointing out that this reduces the risk to their offspring from predators attracted by the white lining of the shells, rendering ancestral birds exhibiting this behaviour more likely to have offspring that survive and reproduce.[13] In the study of animal behaviour, an explanation of this kind is typically described as a *functional explanation*; the *function* of a trait is here understood to be the production of an effect of which it holds that the disposition of the trait to produce such an effect led to its selection.[14] More snappily, a trait can be said to have the function of doing whatever it was selected to do. Functional explanations typically involve teleological language, saying that a trait exists *in order to* generate a certain effect: for example, we say that black-headed gulls dispose of empty shells in order to protect their chicks from predators. As we understand the locution, a trait exists in order to generate an effect iff that trait was the object of selection due to its capacity to generate such an effect.

Paradigmatic as it may be, functional explanation does *not* exhaust the broader category of evolutionary explanation. Many evolutionary changes, especially in small, isolated populations, are due to random differential survival/reproduction, called *genetic drift*. Certain traits evolve under the influence of natural selection in spite of being selectively neutral: they may be linked to a beneficial trait at the genetic level (*pleiotropy*) or constitute an epiphenomenon of a beneficial trait (a *spandrel*).

[13] Tinbergen et al. (1962).
[14] See, *inter alia*, Davies et al. (2012), Laland & Brown (2011). See also footnote 5.

Another kind of evolutionary explanation that will be important in the following is an explanation in terms of *phylogeny*. Just as a genealogy locates an individual within a 'family tree' stretching back over several generations, so a phylogeny locates a species (or other taxon) within a 'tree of life' stretching back over millions (or billions) of years. According to our current understanding, the ancestry of every living thing can in principle be traced back to one organism: *LUCA*, the *last universal common ancestor*. It is the task of phylogenetics to describe the patterns of inheritance and divergence leading from LUCA to the millions of distinct species we observe today.

Importantly, there are cases in which phylogeny can explain what natural selection cannot. The horizontal tail-flukes of whales are a case in point. The horizontal flukes are, in one respect, an obvious adaptation for swimming. But why have whales evolved flukes that are *horizontal* and not *vertical* like the caudal fins of fish? Whales do not inhabit special ocean environments in which horizontal flukes confer some advantage over the vertical alternative. The answer lies instead in cetacean phylogeny.[15] Whales are descended (and fish are not) from terrestrial mammals in the *Cetartiodactyla* clade that ran by flexing their spinal columns in the vertical plane. Natural selection is a tinkerer, and in whales it has adapted this terrestrial system of spinal motion to swimming, necessitating the evolution of horizontal flukes that can be waved up and down. Thus, constraints imposed by phylogeny explain the different tails of whales and fish.

The final point I'd like to address in the subsection is the well-known distinction between *proximate* and *ultimate* causes, which speaks to the limitations of evolutionary explanations.

In general, explanatory questions admit of multiple, non-competing answers.

[15] Gould (1994).

Conversational context may favour one particular answer as uniquely appropriate, but that merely renders the rest inappropriate, not false. To pick an example from history, the question *Why did a world war break out in July 1914?* can be answered with *Because Archduke Franz Ferdinand was assassinated in June* or *Because there existed a network of military alliances joining the fates of the European Great Powers.* These answers do not exclude one another: they are complementary.

When it comes to the explanation of traits in biology, there exist well-established categories by which complementary explanations can be typed. Most famous here is Ernst Mayr's (1961) distinction between explanations citing *proximate* and *ultimate causes.*[16] Proximate causes are causes of a trait that operate within an organism's own lifetime: these might include the immediate triggering causes or the developmental factors responsible for its acquisition and expression. Ultimate causes belong to evolutionary history: an explanation in terms of function or phylogeny is an explanation in terms of ultimate causes. In Mayr's work, the distinction between proximate and ultimate causes is connected to a number of more controversial theses, such as a strict separation of evolutionary and developmental biology.[17] Nonetheless, the proximate/ultimate distinction is widely acknowledged throughout the life sciences, alongside the importance of keeping in mind that proximate and ultimate factors are *not* competing, but complementary. As Mayr (1961) notes: "the biologist knows many heated arguments about the 'cause' of a certain biological phenomenon that could have been avoided if the two opponents had realized that one of them was concerned with proximate and the other with ultimate causes." (503)

The key message here is that factors like natural selection and phylogeny are only ever elements in a broader explanatory picture in which proximate factors must

[16] Tinbergen (1963) is often credited with a more fine-grained explanatory taxonomy that cuts across Mayr's distinction.
[17] See Ariew (2003), Calcott (2013), Laland et al. (2012), and Laland et al. (2011) for concerns about Mayr's implementation and use of the proximate/ultimate distinction.

also play their part. As we'll see in chapter 3, this point has not always been properly observed in debates about the debunking power of evolutionary explanations.

2.3 Scope

In unpacking our question, one final issue remains to be addressed. We are considering whether evolutionary explanations can debunk our moral beliefs. Is this supposed to cover *all* moral beliefs or just a proper subset?

I want us to understand the question in such a way as to be neutral with respect to this issue of scope. Thus, we are to consider the question of whether evolutionary explanations have the capacity to debunk *at least some* of our moral beliefs, leaving open how much of our moral outlook might potentially be subject to debunking. Philosophers who are otherwise agreed that our question should be answered in the affirmative diverge on the issue of scope because they disagree on the extent to which our moral beliefs can be explained in debunking evolutionary terms. Some believe we have a means of accessing ethical truths that is suitably independent of our evolved biases;[18] others are skeptical of this.[19]

I think the issue of scope can't be meaningfully settled until we have a proper grasp of why evolutionary explanations are debunking. I believe, furthermore, that most contemporary philosophers are misled on this point, because they focus on issues of function at the expense of phylogeny. Most evolutionary debunking arguments today are driven by (what I call) the issue of *functional truth-irrelevance*: the availability of functional explanations for elements of human moral psychology which make no reference to the truth of our moral beliefs. I believe that a successful debunking argument can be constructed only if we bring issues of phylogeny back into the

[18] Crisp (2006); de Lazari-Radek & Singer (2012); FitzPatrick (2008); Greene (2008); Huemer (2008); Parfit (2011); Radcliffe-Richards (2000); Singer (2005, 2006). Cf. Nagel (1978).
[19] Berker (2009); Joyce (2006); Kahane (2011); Street (2006); Tersman (2008).

picture, and, in particular, advert to some element of (what I call) *phyletic contingency* in our moral beliefs, showing that our moral outlook reflects arbitrary contingencies of our phylogeny, much as the whale's tail reflects the conditions of its descent from terrestrial quadrupeds. Phyletic contingency has a good claim to being the original worry when it comes to the potential of evolutionary explanations to debunk our moral beliefs, as I explore in the next section. The emphasis on functional truth-irrelevance is a more recent development.

3. Evolutionary debunking arguments from Darwin to today

3.1 Victoriana

The thought that evolutionary explanations are debunking is by no means recent. It has existed since Darwin first published an account of the evolutionary origins of human morality in *The Descent of Man* (1871). It quickly became a feature of late-Victorian intellectual life. Thus, in his *Foundations of Zoology* (1899), the leading American Darwinist William Keith Brooks could write, without referring to anyone in particular:

> Many good and thoughtful people hold that proof that our moral sense has had a natural history would have very dreadful consequences; that it would show that duty is not duty, right and wrong neither right nor wrong, and that the significance man has attributed to this part of his nature a mistake. (118)

In seeking to explain the 'moral sense' as arising through the gradual modification under natural selection of traits shared with other species, Darwin (1879/2004) had posited that "any animal whatever, endowed with well-marked social instincts, would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well developed, or nearly as well developed, as in man." (120-121) He claimed that the social instincts which undergird the 'moral sense' are "retained from

an extremely remote period" (132) having most likely originated with our "early ape-like progenitors." (133) Since the moral sense is founded on inherited social instincts, we should not expect, Darwin claimed, that any social animal whose intellectual faculties are similar to those of human beings would acquire exactly the same moral beliefs: the moral outlook of an organism will reflect idiosyncrasies of its phylogeny. "If, for instance, … men were reared under precisely the same conditions as hive-bees, there can hardly be a doubt that our unmarried females would, like the worker-bees, think it a sacred duty to kill their brothers, and mothers would strive to kill their fertile daughters; and no one would think of interfering." (122)

Contemporary reactions to Darwin's theory were highly negative, as evinced by three hostile reviews published in the spring of 1871 in *The Quarterly Review*, *The Edinburgh Review*, and *The Theological Review*. To some extent, the hostile reaction is attributable simply to the fact that Darwin offered a naturalistic explanation for morality, closely linked in Victorian consciousness with religion. Robert J. Richards (1987) suggests that Darwin's "Victorian critics heard the sea of faith's melancholy, long, withering roar." (241)

It's clear, however, that contemporary critics were particularly distressed by Darwin's remarks about bees, with the concomitant suggestion that our own moral outlook reflects arbitrary contingencies of our phylogeny - our status as hominins, as opposed to hymenoptera. Thus, the anonymous author of *The Edinburgh Review* piece, having cited Darwin's claims about the moral sense of hypothetical intelligent bees, writes: "The sense of right and wrong, according to this view, is no definite quality, but merely the result of the working together of a series of accidents … We need hardly point out that if this doctrine were to become popular, the constitution of society would be destroyed" (217). A similar view was taken by Frances Cobbe (1871) in *The Theological Review*. She too notes that under Darwin's theory, "Conscience is the

result of certain contingencies in our development" (192). On this view, our moral sense is merely "the provincial prejudice, as we may describe it, of this little world and its temporary inhabitants" (175). Cobbe too is quick to engage in lurid doomsaying, writing of Darwin's theories: "in the hour of their triumph would be sounded the knell of the virtue of mankind." (175)[20]

3.2 Contemporary discussion

These reflect the worries current in Darwin's own lifetime. Contemporary interest in evolutionary debunking arguments owes much to the emergence of *sociobiology* in the 1960s and 1970s. Through the work of William Hamilton (1964) and Robert Trivers (1971), sociobiology made important strides in understanding the evolution of social behaviours associated with morality, including altruism, reciprocity, and moralistic aggression. Thus, in E. O. Wilson's firebrand popularization, *Sociobiology: The New Synthesis* (1975), he claimed: "the time has come for ethics to be removed temporarily from the hands of the philosophers and biologicized." (562) In particular, Wilson (1975: 562-564) argued that moral philosophy is limited by its ignorance of the evolutionary origins of the intuitive responses that guide philosophers in building their theories.

Wilson's remarks on ethics consisted largely of undersupported conjectures enlivened with characteristic bluster, but the subsequent decade saw a number philosophers attempt to address the potential for sociobiological explanations to impact on our moral beliefs, including Peter Singer (1981), Robert Nozick (1981), and Derek Parfit (1984). Each adopted a relatively sanguine attitude. Thus, according to Parfit (1984): "When some [ethical] belief or attitude has an evolutionary explanation, this, in itself, has neutral implications. It cannot by itself show that this

[20] Cobbe's review is notable in that it drew responses from both Sidgwick (1872) and Darwin (1879/2004), with Darwin also commenting (in the same place) on Sidgwick's reply to Cobbe.

belief or attitude either is or is not justified." (186) At best, Parfit suggests, evolutionary explanations may undercut attempts to argue abductively that since a particular ethical belief is very widespread it must be true.[21]

The philosopher of biology Michael Ruse (1986) adopted a more radical stance. According to Ruse, evolutionary explanations support a view on which "morality is a collective illusion foisted upon us by our genes." (253) Ruse positions his argument as one according to which evolutionary considerations undermine belief in moral objectivity and undermine our first-order beliefs by virtue of their assumed presupposition of an objective morality - an assumption Ruse derives from Mackie (1977). Ruse and Wilson (1986) co-authored a paper arguing for similar conclusions. However, skeptical views of this kind appear to have received little serious attention within mainstream moral philosophy until quite recently. Thus, Ronald Dworkin (1996) could dismiss as "plainly mistaken" the view "that a successful Darwinian explanation of moral concern … would have skeptical implications" (123) - without so much as an argument.

The current philosophical climate takes seriously the possibility that evolutionary explanations are debunking. One contributing factor is likely to have been the recent growth in empirical moral psychology, and in related disciplines which focus on the evolution of morality. Whereas it might once have been possible to disregard Ruse's debunking argument because his speculations about evolution and morality were, by his own admission, "beyond the evidence" (1986: 235), there is now much less room for these reservations, as will be made clear in chapters 1 and 2. In addition, there has been a general trend within analytic philosophy to go beyond armchair

[21] Singer (1981) does say: "Seeing that an ethical principle has a biological basis … undermines it" (158). However, he too appears to delimit this to undercutting the evidentiary value of consensus. In full, he says: "Seeing that an ethical principle has a biological basis does not support that principle. If anything, it undermines it, by showing that its widespread acceptance is no evidence that it is some kind of absolute moral truth." (158) In fact, Parfit (1984) cites Singer as sharing his view.

reflection and engage more seriously with relevant empirical disciplines,[22] not least within moral philosophy.[23]

Much of the responsibility for launching evolutionary debunking arguments as a topic of discussion in its own right goes to Richard Joyce (2000, 2001, 2006) and Sharon Street (2006, 2008a, 2011). Joyce believes, like Ruse, that evolutionary explanations are debunking with respect to morality as a whole. Street believes that evolutionary explanations are debunking iff some form of *Meta-Ethical Realism* is presumed - where *Meta-Ethical Realism* is the view that moral facts are constitutively independent of our evaluative attitudes. Evolutionary explanations constitute a serious problem for realists, Street believes, but need not otherwise alter our first-order ethical beliefs if we reject *Meta-Ethical Realism*. Street's argument in particular has attracted extensive critical commentary.

It is today quite common to find evolutionary debunking arguments discussed alongside more familiar sources of ethical anti-realism, such as widespread disagreement or 'queerness'.[24] We also find prominent philosophers deploying evolutionary considerations in a targeted fashion, hoping to debunk particular views while leaving their favoured theories standing: Roger Crisp (2006) repeatedly employs this strategy; Peter Singer (2005, 2006) has consistently argued that evolutionary considerations selectively undermine non-utilitarian moral beliefs.[25] Whether evolutionary considerations can be deployed in this selective fashion is a matter of debate.[26]

Another prominent issue in contemporary discussion concerns the extent to which evolutionary debunking arguments require particular meta-ethical presuppositions. Like Street, some philosophers believe that evolutionary explanations

[22] Ladyman & Ross (2007); Knobe & Nichols (2008).
[23] Appiah (2008).
[24] See, *e.g.*, Enoch (2011), Huemer (2005), Kramer (2009), Parfit (2011).
[25] Cf. Greene (2008), de Lazari-Radek & Singer (2012).
[26] For references, see footnotes 18, 19.

are debunking only if we presume some form of *Meta-Ethical Realism*.[27] An interesting discussion has arisen as to whether *Quasi-Realism* is sufficiently realist to support evolutionary debunking arguments or can evade the evolutionary challenge posed by Street.[28] There are also some, including Joyce (forthcoming), who doubt that meta-ethical presuppositions of this kind are relevant in assessing the capacity of evolutionary explanations to debunk our moral beliefs. The relevance of presuppositions about moral objectivity turns, I believe, on why we think of evolutionary explanations as debunking. As we'll see in later chapters, Street's argument relies on considerations relative to which the rejection of *Meta-Ethical Realism* makes a difference, whereas Joyce's arguments are of a kind where this issue is moot.

As this suggests, debunking arguments come in different kinds, emphasizing different features of the evolutionary process and relying on a variety of epistemic principles. Whereas some authors have sought to reduce all such arguments to a single canonical form,[29] I think it best to leave their heterogeneity more open to view.

That having been said, all of the most prominent arguments currently discussed do share one important feature: they are narrowly focused on the issue of functional truth-irrelevance. These arguments seek to infer that evolutionary explanations are debunking because evidence shows the moral mind to lack the evolved function of 'tracking' ethical truths: if some element of human moral psychology has evolved under selection, the advantages conferred by this trait are explanatorily independent of the accuracy of any corresponding moral beliefs. The issue of phyletic contingency may be noted in passing or as an aside, but never as the basis for a fully-realized debunking argument.[30] This, I believe, is a mistake.

---

[27] *E.g.*, Kahane (2011).
[28] See Street (2011) arguing for this view and Gibbard (2011) in reply.
[29] *E.g.*, Kahane (2011), Mason (2010).
[30] See, *e.g.*, Crisp (2006: 17-18), Joyce (2006: 229), Street (2006: 120-121). Crisp is somewhat anomalous

4. Aims and structure of the essay

4.1 Aims

In this essay, I will argue that the most prominent debunking arguments discussed in the literature today fail. To construct a successful debunking argument, we need to go back to our roots and take up once more the issue of phyletic contingency. I do just that, describing how to construct a cogent debunking argument rooted in considerations of phyletic contingency.

More generally, the overarching aim of this essay is to rethink evolutionary debunking arguments from the ground up: to explore exactly what evolutionary explanations must be like in order to be debunking, whether we have evidence for explanations of that type, and, most importantly, what kind of epistemic principles could allow us to say that explanations of that kind are debunking. Once we change our mind about these issues, questions about scope and meta-ethical presuppositions take on a new character, as I explore in my conclusion.

The first task in this kind of project is to consider whether our present understanding of the evolutionary origins of morality lends support to the claims on which debunking arguments rely. This isn't beyond dispute,[31] so a review and discussion of relevant evidence is necessary to get the project off the ground. I devote most of my attention, however, to epistemological questions: that is, to exploring what epistemic principles are required to infer the presence of defeaters from our awareness of evolutionary explanations. This is where the current discussion is most in need of

in this respect, however. Although he does not provide a detailed account of why he considers evolutionary explanations to be debunking, what little he says emphasizes phyletic contingency. Thus, he says, "I find myself perfectly able to accept a debunking view of morality, the functions of which are so specific to the human condition." (18)

[31] See Machery & Mallon (2010).

improvement.[32] Most philosophers who address the question of whether evolutionary explanations can debunk our moral beliefs do not specialize in epistemology - and it shows. The epistemological presuppositions relied on in popular debunking arguments are often difficult to pin down; much of my discussion in the following will be devoted to exposing these epistemic foundations. Similarly, connections to key issues in contemporary epistemology often go unnoticed; I hope to ensure that the relevant dots are joined.

4.2 Structure

Here is the structure of this essay. It is divided into six main body chapters, followed by a conclusion. The six chapters are divided into three pairs.

The first two chapters consider what contemporary scientific evidence can tell us about the evolutionary origins of morality, and, in particular, to what extent the evidence speaks in favour of the claim that our moral psychology exhibits functional truth-irrelevance and/or phyletic contingency. Chapter 1 sets out evidence in support of two hypotheses: firstly, that our moral outlook is influenced by innate biases with recognisable homologues in closely related primate species; secondly, that the moral norms we observe across human societies have been shaped by natural selection in the context of group competition. Chapter 2 considers to what extent the two hypotheses for which I've offered evidence support the claims on which debunking arguments rely: that our moral psychology exhibits functional truth-irrelevance and/or phyletic contingency. These claims cannot be straightforwardly read off from the scientific work discussed in chapter 1, as they involve a number of meta-ethical questions.

The next two chapters offer a critique of pre-existing debunking arguments that are rooted in the issue of functional truth-irrelevance. Chapter 3 considers two

[32] Cf. White (2010).

lines of argument particularly associated with Ruse and Joyce. The first relies on an appeal to *Ockham's Razor*, the second on considerations of Nozickian sensitivity. I note that both are subject to a range of epistemological problems. They also share a deeper, non-epistemological flaw, involving a confusion of proximate and ultimate explanatory factors. Chapter 4 considers a line of argument particularly associated with Street, rooted in what I call the *Coincidence Problem*: the thought that it would require an extraordinary coincidence if natural selection has favoured the evolution of moral beliefs which turn out also to be (objectively) true. I argue that the appearance of a problem here is misleading: a commitment to a coincidence of this kind is not unacceptable. Central to my argument will be an analogy between the *Coincidence Problem* and the *Fine-Tuning Problem* in the philosophy of cosmology.

The final chapters develop a novel argument designed to show that evolutionary explanations can nonetheless undermine our moral beliefs due to the issue of phyletic contingency. Chapter 5 asks us to consider evidence of phyletic contingency as an instance of a more general epistemic phenomenon, which I call *contingency anxiety*.[33] This refers to the sense of unease we often feel when we discover that we hold certain beliefs due to arbitrary features of our background, such as the identity of our parents, the culture in which we were raised, our gender, *etc.* Relying on a suggestion due to Roger White (2010), I argue that whether cases of contingency anxiety involve defeaters turns on the epistemology of disagreement. Thus, the surprising conclusion at which we arrive is one according to which our attitude to the question of whether evolutionary explanations can debunk our moral beliefs will be a function of our view about the epistemic significance of moral disagreement. In chapter 6, I argue for a concessive account of moral disagreement, which supports the conclusion that evidence of phyletic contingency is debunking. Central here will be an

[33] Inspired by 'genealogical anxiety', as coined by Amia Srinivasan.

argument for the claim that we cannot treat our own moral intuitions as enjoying any fundamental privileged status with respect to the dissenting intuitions of others.

These six chapters are rounded off by a conclusion, which summarizes my argument and sets out the key questions which arise in its wake. In particular, I outline how we must reconceive the question of the scope of evolutionary debunking arguments and what we should be looking for when considering whether the capacity of evolutionary explanations to debunk our moral beliefs requires any particular set of meta-ethical presuppositions.

1.

The evolutionary origins of morality: considering the evidence

1. Introduction

This chapter considers what evidence there is to support evolutionary hypotheses of the kind required by debunking arguments. Research on moral origins is a rapidly-expanding, interdisciplinary enterprise. This chapter is not – and could not be – a general overview of the field: that would require a separate monograph altogether. What I mean to offer is a selective treatment, designed to elucidate and support hypotheses that are plausible, given present evidence, and serviceable as a basis for evolutionary debunking arguments.

We can begin by noting that there is an uncontroversial sense in which evolutionary considerations illuminate the character of human moral psychology. As Robert Brandon (1989) notes: "all phenotypes are to some extent the products of the process of evolution by natural selection." (41) The claim that human moral psychology is the result of evolution is thus close to truistic, and correspondingly innocuous. To arrive at a view about the evolutionary origins of morality fit to serve as a basis for debunking arguments, we have to develop a more specific set of proposals.

I take the available evidence to support two hypotheses: firstly, that the moral beliefs of human beings are influenced by a suite of innate affective biases inherited from the last common ancestor of humans, chimpanzees, and bonobos; secondly, that the moral norms characteristic of human societies are systems of social control adapted to promote cooperation and suppress competition, evolved via gene-culture co-evolution and group selection. My aim in this chapter is to faithfully summarize the

evidence supporting these hypotheses. The next chapter will take a more philosophical approach to this material, outlining how the evidence amassed here speaks in favour of the claim that our moral psychology exhibits functional truth-irrelevance and phyletic contingency.

Here is the plan for this chapter. Section 2 will take up issues related to conceptual analysis and stage-setting. Here, I clarify what I mean by *innateness* and *adaptation*, emphasizing the distinctness of these phenomena; I also situate the hypotheses for which I argue relative to the options otherwise available to those who seek an evolutionary perspective on human morality. In section 3, I argue for the claim that our moral outlook reflects innate affective biases inherited from the last common ancestor of *Homo* and *Pan* (the respective genera dividing humans from chimpanzees and bonobos). This argument comes in two parts. The first shows the extent to which our moral outlook is influenced by our emotional responses; the second provides evidence of affective dispositions in chimpanzees corresponding to universal (or nearly-universal) domains of moral concern in our species. In section 4, I argue for the view of morality as a group-level adaptation. I present a number of reasons to suppose that large-scale cooperation in our species must be explained by appeal to some form of group selection, and I outline how group competition will have favoured societies that rely on moral norms to suppress antisocial behaviour and promote cooperation within the group.

2. Key concepts and their meanings

I have put forward two hypotheses for which I intend to outline supporting evidence: one about innate biases, the other about adaptation. Before considering the evidence, I need to clarify certain key terms and concepts on which I'll rely. I begin by analysing the concepts of *innateness* and *adaptation*, emphasizing the distinctness of the

phenomena to which they refer. There is then further groundwork to be done in terms of outlining how my favoured hypotheses stand out from other forms of moral nativism and adaptationism, and I take up this issue in sections 2.2 and 2.3.

2.1 Adaptation and Innateness

I'll begin by analysing the concept of *adaptation*. An adaptation is a trait that has evolved under natural selection because it has raised the relative fitness of its bearers.[1] Not every trait produced by selection is an adaptation, so understood: the trait must have evolved *because of* its beneficial effects. The rhythmic sound produced by the heart is the product of selection, but is in an incidental by-product of selection for an organ capable of sustaining regular blood-flow. A trait that constitutes an adaptation is said to have been *selected for*; a by-product has merely been *selected of*.[2]

A brief note about the concept of natural selection is also in order. Characterised in the abstract, selection occurs when the members of a population vary in ways that lead some to survive and reproduce at greater rates than others, provided that those traits which lead to increased reproduction are reliably transmitted to offspring: these traits will increase in frequency over time as a matter of selection. What's important to realize is that the concept of selection, so understood, is *substrate neutral*.[3] It imposes no requirements on the intrinsic nature of the units of selection, nor the mechanism by which their traits are transmitted. It's no part of the concept of selection that it operates among genes or organisms, nor that inheritance involves the transmission of biological material such as chromosomes. One can in principle apply the concept of natural selection to replicating algorithms or entire universes.[4] So long as we have variation, differential reproduction, and heritability we have selection.

[1] Brandon (1989); Sober (1984).
[2] This distinction is due to Sober (1984).
[3] Dennett (1995); Lewontin (1970).
[4] See Smolin (1997) on cosmological natural selection.

Importantly, this means that traits passed on via social learning are still liable to be classified as adaptations. I'll return to this point at this end of this subsection.

Let's move on now to examining the concept of *innateness*. This concept has been subject to fierce controversy in some quarters, with calls for its elimination from our conceptual repertoire.[5] The eliminativists are not without a point. There is good evidence that our folk biological thinking incorporates a vernacular innateness concept that runs together a number of unrelated biological properties and encourages a range of mistaken inferences.[6] Most philosophical accounts of innateness are not attempts to characterise the biological intuitions of the folk, however. They are aimed at describing the specific role played by the innateness concept in the cognitive sciences. This need resemble the vernacular innateness concept no more than the species concept used in biological taxonomy resembles the discredited, essentialist concept deployed in folk biology.[7]

Two analyses of innateness strike me as most promising: *Closed-Process Invariantism* and *Primitivism*. Both may be thought of as attempts to cash out the rough suggestion that a cognitive trait is innate to the extent that its development is independent of environmental cues and constrained by endogenous genetic factors. According to *Closed-Process Invariantism*, a psychological trait is innate for a given genotype iff it would emerge across the range of normal development environments and it develops via a *closed process*, where the 'closedness' of a process is inversely proportional to the diversity of traits in which it can eventuate.[8] According to *Primitivism*, a psychological trait is innate for a given genotype iff it would emerge across the range of normal developmental environments and its development falls

[5] See Bateson (1991), Griffiths (2002), Lehrman (1953), Tinbergen (1963).
[6] Griffiths (2002); Griffiths et al. (2009); Mameli & Bateson (2011).
[7] On folk biology and essentialism see Medin & Atran (2004).
[8] For further details see Mallon & Weinberg (2006).

beyond the explanatory purview of psychological science.[9] On this construal, the innateness concept serves as a disciplinary boundary-marker: innate traits are those that must be taken as given for the purposes of developmental psychology.

For our purposes, it does not matter which of these accounts we prefer, and so I won't attempt to weigh their pros and cons. Instead, I want to wrap-up this subsection by emphasizing the distinctness of the phenomena we've just been analysing. In particular, I want to underscore the point that not every innate trait is an adaptation, nor is every adaptation innate.

Some analyses don't support this dissociation. Richard Joyce (2006) has proposed that "what people generally mean when they debate the 'innateness of morality' is whether morality (under some specification) can be given an adaptive explanation in genetic terms" (2). This suggestion to define innateness in terms of adaptation fails to square with established usage. Many dysfunctions are classified as innate. In the first paper to describe the syndrome, Kanner (1943) classified autism as "an innate inability to form the usual, biologically provided affective contact with people." (250) He did not mean to suggest that autism is an adaptation. Whether a trait is properly classified as innate is thus independent of whether it has been selected for and determined only by the extent to which its development is independent of environmental cues.

Conversely, not every adaptation must be innate, because natural selection is substrate neutral. Since the concept of selection sets no requirements on the mechanism by which favoured traits are copied, adaptations can in principle be elements of cultural inheritance. There is no inherent opposition between evolution and culture. A trait that is selected for but inherited via social learning will be an adaptation, but not innate. Traits of this kind are *cultural adaptations*, and are uniquely

[9] For further details see Cowie (1999), Samuels (2002, 2004).

important in human evolution. In my view, cultural adaptation is an important process in the evolution of morality, an issue to which I'll return in section 2.3.

2.2 The Innate Biases Model

I propose that the moral beliefs of human beings are influenced by a suite of innate affective biases inherited from the last common ancestor of *Homo* and *Pan*. Having explained the general concept of innateness, I want to use this subsection to outline what makes this proposal distinct as a form of moral nativism.

The proposal is an instance of what Chandra Sripada (2008) calls the *Innate Biases Model*. The *Innate Biases Model* is a form of *Content Nativism*, according to which innate structure shapes the content of human moral judgments. *Content Nativism* may be contrasted with *Capacity Nativism*, according to which we have an innate specialization for acquiring moral norms without necessarily being predisposed to accept norms with any particular content.

The *Innate Biases Model* takes its inspiration from Dan Sperber's (1996) work on the *epidemiology of representations*. This research program seeks to explain how certain ideas arise and spread across human minds, and how some are able to propagate successfully so as to become stable elements of shared culture.[10] The epidemiology of representations ties the study of culture closely to cognitive science on the assumption that any successful account of the spread of ideas must be informed by an understanding of the internal mechanisms by which certain representations are apt to capture our attention, become stored in memory, and reconstructed so as to be passed on to others. Sperber especially stresses the importance of innate mental structure in biasing the propagation of certain ideas.

[10] This is related in spirit to the theory of *Memetics* (Blackmore 1999; Dawkins 1976), though Sperber (1996: 100-106) explicitly rejects the existence of *memes*, conceived as particulate units of cultural inheritance comparable to genes.

The approach is well-illustrated by the work of Pascal Boyer (2001), Justin Barrett (2000), and others on the epidemiology of religious ideas. This seeks to explain the pervasiveness of religious belief across human cultures by describing innate psychological mechanisms that bias the propagation of concepts related to supernatural agency. As biases, these favour the emergence of certain beliefs without strictly determining their acquisition. In addition, Boyer *et al.* do not presuppose the existence of innate religious concepts, hoping to explain the prevalence of religious ideas by reference to more general features of the innate mind. For example, one property of religious concepts that is supposed to make them apt to gain widespread cultural ascendance is their status as *minimally counterintuitive* relative to our innately prespecified expectations about the nature of agents, organisms, and artefacts.[11] Concepts which deviate only slightly from these intuitions are salient in light of their surprisingness, but also readily processed and reconstructed due to their overall conformity to our expectations.[12]

The *Innate Biases Model* employs an analogous approach to the study of human moral psychology. The model hypothesizes that certain elements of innate mental structure make certain moral beliefs more likely to arise and spread. As in the case of religious ideas, these biases render certain moral beliefs more probable without strictly determining their acquisition. As we'll see in section 3, this renders the *Innate Biases Model* well-suited to explaining cross-cultural patterns of unity and diversity in moral norms: norms tend to cluster around certain issues, but are otherwise highly variable.

The innate psychological mechanisms posited by the *Innate Biases Model* are also not assumed to have any moral content of their own, just as Boyer and Barrett do not assume the existence of innate religious beliefs. Emotions including sympathy, anger, and disgust have been hypothesized to act as innate biases in the moral

[11] Boyer & Ramble (2001).
[12] For relevant evidence see Barrett & Nyhof (2001).

domain,[13] and my own proposal is squarely in this vein. By positing innate biases without innate moral content, the *Innate Biases Model* contrasts with other forms of *Content Nativism*. According to the *Simple Innateness Model*, certain determinate moral principles are innate.[14] A more sophisticated proposal is *The Moral Parameters Model*,[15] taking its inspiration from Chomsky's *Principles and Parameters Model* in linguistics.[16] *The Moral Parameters Model* posits a series of innate moral principles containing free parameters whose values are subsequently fixed by cultural instruction. According to one proposal, the human mind is furnished with an innate harm norm containing a free parameter whose culturally variable setting determines the class of agents not be to be harmed in various ways.[17]

One final notable feature of my proposal is that it explicitly posits a partial phylogeny for the innate biases. Drawing on research aimed at identifying 'building blocks' of human morality in closely related primate species, I suggest that many of the affective dispositions which bias our moral judgments are traceable to a distant primate ancestor. I don't mean to suppose that our moral beliefs are influenced by affective traits exactly like those observed in closely related primate species, these having been conserved from our common ancestor without modification. This view isn't especially plausible, as many of the affective biases I identify in chimpanzees exhibit an *egocentric bias*. For example, chimpanzees protest at unequal distributions of rewards, but it is only the disadvantaged individuals who protest. By contrast, someone who thinks that inequality is morally bad will be motivated to protest at inequality even if they're on top. The view I mean to present is one on which traits like the inequity aversion exhibited by chimpanzees and the generalized moral emotions characteristic of our species are *homologues*: variants of the same trait whose

[13] Haidt & Joseph (2004, 2007); Nichols (2004, 2005); Sripada (2008).
[14] See Harman (1999), Mikhail (2011) on the intention/foresight distinction.
[15] Dwyer et al. (2010).
[16] Chomsky (1981, 1995); Baker (2001).
[17] See Dwyer (1999, 2006), Harman (1999).

underlying similarities result from shared inheritance.

2.3 Morality as a Group Level, Culture-Bound Adaptation

I want now to describe certain distinctive features of the approach I favour when it comes to thinking about morality as an adaptation. Three features in particular will be addressed: the idea of morality as a form of *social control*, the appeal to *gene-culture co-evolution*, and, finally, the appeal to *group selection*.

In social organisms, group life is subject to certain forms of order and regularity. For example, in many primate species, such as chimpanzees, social life is organized around a despotic hierarchy. The hierarchy serves to mediate disputes, since each individual knows its position within the hierarchy and will submit to a more dominant individual in the event of conflict.[18] Human beings have lived without a similar despotic hierarchy throughout most of our evolutionary history. Nomadic hunter-gatherers are characterized by egalitarian relations among adult males, with efforts to assert dominance actively suppressed; in the place of a linear dominance hierarchy, conflicts and disputes are resolved by reference to moral norms upheld and enforced by broad coalitions, employing sanctions ranging from harsh words to collective execution.[19] Moral norms enforced by the group thus serve to organize social life by encouraging and suppressing certain behaviours, thereby delimiting conflict and fostering cooperation. That's what I have in mind by speaking of morality as a system of social control.

Let's turn now to the concept of *gene-culture co-evolution*. Although rudimentary cultures are found in other species including chimpanzees,[20] human beings are distinctive in the scale and extent of our cultural variability. We also appear

[18] See Goodall (1982).
[19] Boehm (1999, 2000).
[20] See Whiten et al. (1999).

unique in our ability to establish evolving traditions in which cultural traits are subject to cumulative, incremental improvements.[21] We are thus beneficiaries of a *dual inheritance*, one genetic, the other cultural, with both subject to analogous patterns of evolutionary change and adaptation.[22] *Gene-culture co-evolution* refers to evolutionary change caused by the reciprocal interaction of cultural and genetic information. A prominent example is the lactose tolerance found among descendants of dairying cultures in Europe and West Africa: the cultural practice of herding cattle created an environment in which selection favoured the ability to digest lactose as a means of obtaining added nutrients from dairy products.[23]

To the extent that we think of certain aspects of moral evaluation as adaptations, I think selection among cultural variants plays a large role. The adaptationist hypothesis I favour concerns moral systems comprising relatively determinate norms of the sort we might find in the Code of Hammurabi or the Pentateuch. Human beings everywhere create moral norms,[24] but, as I'll argue in section 3, their variability is otherwise so great as to render any form of *Content Nativism* more ambitious than the *Innate Biases Model* implausible. Determinate moral norms should thus be thought of as cultural traits, acquired and transmitted via social learning. At the same time, we should expect that elements of innate mental structure make certain norms more likely to emerge and that culture-bound moral systems will feed into genetic evolution: for example, favouring genes which underpin cooperative behaviours prescribed by prevalent norms and disfavouring genes that predispose individuals to aggressive, antisocial behaviours.[25] Thus, the evolution of morality involves the reciprocal interaction of culture and genes.

---

[21] Dean et al. (2012); Tomasello (1999).
[22] See Boyd & Richerson (1985), Cavalli-Sforza & Feldman (1981), Durham (1991), Lumsden & Wilson (1981), Richerson & Boyd (2005).
[23] Curry (2013); Feldman & Cavalli-Sforza (1989).
[24] Brown (1991).
[25] See Richerson & Boyd (2005), Boehm (1999).

The final notable feature of my approach is that my adaptationist proposal appeals to *group selection*. As many readers will know, appeals to group selection have been subject to controversy since critical attacks launched by John Maynard-Smith (1964) and George C. Williams (1966) in the 1960s. In the 1970s, Richard Dawkins (1976) could confidently proclaim that "group-selection theory now commands little support within the ranks of those professional biologists who understand evolution." (8) This orthodoxy has since disintegrated.[26] To the extent that evolutionary biologists remain negatively disposed toward appeals to group selection, this is largely attributable to a perceived lack of conceptual clarity.[27] To forestall this issue, I'll briefly outline how I propose to understand group selection.

One key thing to note is that the concept of group selection doesn't require any reconceptualization of the notion of evolutionary fitness. As typically understood, group selection does not involve the differential survival and reproduction of groups *qua* groups.[28] Rather, group selection occurs when certain evolutionary trends cannot be explained except by the impact of group differences on the fitness of their individual members and by the tendency of groups to compete.

A distinction can be made between *weak* and *strong group selection*. *Strong group selection* occurs when groups whose members exhibit a certain trait enjoy greater average fitness, but those individuals exhibiting the trait are less fit on average than other members of the group. In this case, the trait represents a form of *biological altruism*: it promotes the fitness of others at the expense of the bearer. *Weak group selection* occurs when the group structure characteristic of a population furthers the evolution of a trait that is *not* selected against within groups. Thus, we might have a case where several behaviours can be stabilized by intragroup processes, with selection

[26] Wilson & Wilson (2007).
[27] See West et al. (2010).
[28] For discussion see Okasha (2001).

across groups favouring the evolution of some ahead of others. This is known as *equilibrium selection*. As I discuss in section 4, I believe that the hypothesis of strong selection should be taken seriously when thinking about the evolution of morality.

One final respect in which group selectionist approaches may vary is in whether they confine group variability to the genetic level or include cultural variability as well. As advertised, the approach I favour is of the latter sort.

The approach I favour thus differs sharply from the sort of adaptationist hypothesis offered by Richard Joyce (2001, 2006) and Michael Ruse (1986), which is prominent in a number of philosophical discussions about ethics and evolution.[29] Joyce and Ruse see a disposition to engage in moral thought as an innate trait favoured by individual-level selection: selection has favoured a propensity to believe in the existence of moral requirements as a means of binding human beings to perform individually-beneficial actions to which we are otherwise motivationally unsuited. Moral beliefs are supposed make up the needed motivational oomph required to care for our slowly-maturing offspring or sustain reciprocal cooperation over long time-scales. This proposal strikes me as ill-suited to explain morality from an evolutionary perspective. As I noted in section 2.2, the moral concerns of human beings stand out from similar social instincts in closely related primate species in light of their generality: an organism capable of sympathetic concern is motivated to help others, but someone who accepts a moral rule requiring prosocial behaviour will be interested also in the helping behaviour of third-parties. In short, morality makes us nosy. With its individualistic focus, the view put forward by Joyce and Ruse cannot easily account for this.[30] By contrast, this characteristic feature of moral norms is entirely congruent

[29] *E.g.*, James (2011), Sinnott-Armstrong (2006a).

[30] In his more recent work, Joyce (2006: 115-118) may be thought to address this shortcoming. While setting out the view of moral judgment as a form of personal commitment, he pays considerable attention to the public nature of moral judgments in this regard. In particular, he emphasizes the role played by moral claims in public deliberation and justification: the capacity of moral judgment to motivate actions depends, he claims, on the ability to use moral considerations to justify one's behaviour

with a perspective focused on group competition. The nosiness of morality is indicative of a trait designed to manage social relationships throughout the group, consistent with the assumption that morality is a system of social control favoured by group selection.

3. Human morality as influenced by innate biases

3.1 Introduction

In this section, I will build a case for the first of the hypotheses set out in my introduction: that the moral beliefs of human beings are influenced by a suite of innate affective biases inherited from the last common ancestor of *Homo* and *Pan*. I begin, in 3.2, by summarizing research exhibiting the many ways in which people's emotional states have been found to influence their moral verdicts. In 3.3, I present evidence of affective dispositions in closely-related primates corresponding to cross-cultural domains of moral concern in human beings and summarize evidence for homologous, early-developing dispositions in human children. In light of the evidence already summarized in section 3.2, I suggest that this correspondence is plausibly explained by the tendency of innate affective dispositions to bias the emergence of correlate moral prescriptions in our species. Morality is thus an outgrowth of social instincts shared with other primates.

3.2 Sentimentalism

There's no denying that morality is emotive. Moral arguments quickly raise tempers and voices. Our moral verdicts are often associated with intense feelings: injustice

---

to others (115-116). However, tying up the special motivational powers of moral judgments with third-party reactions in this way presupposes what must be explained: why selection should favour a *prima facie* costly disposition to be concerned about the behaviour of others of a kind that would make public justification necessary.

makes us seethe with anger; corruption leaves us sick with disgust.

The fact that our moral evaluations are emotive leaves open the exact nature of the relationship between moral judgment and affect. According to Bryce Huebner and colleagues (2009), morality is emotive simply because moral judgments typically cause strong emotional responses: our emotional responses are not themselves operative in our ethical evaluations. Others take the available psychological evidence to point in the opposite direction, supporting *Sentimentalism*: the view that emotions play an important role in determining our sense of right and wrong.[31] I side with the latter camp. In this subsection, I present empirical evidence that people's emotional states influence their moral verdicts in multiple ways.

I've been speaking quite generally of emotions influencing moral appraisals. Evidence indicates that particular emotions are related to specific domains of moral concern.[32] The relation is observed in the characteristic causation of certain emotions by particular classes of moral issues: for example, anger by unfairness.[33] We also find domain-specific effects of emotions on moral judgment: for example, inducing disgust selectively intensifies moral disapproval of purity-violating actions, such as sexual promiscuity.[34] We call these emotions which are closely related to some domain of moral concern *moral emotions*.

I will identify a number of ways in which moral emotions influence our moral judgments, running via a multiplicity of causal pathways. In terms of immediate causation, I show that emotional responses can modulate the intensity of our moral judgments and serve to moralize behaviours we would otherwise have thought innocuous. I also present evidence that certain emotions are crucial to the development of a normal capacity for moral judgment. I begin by considering experiments

[31] Haidt (2001); Haidt & Joseph (2004, 2007); Nichols (2004, 2005); Prinz (2007).
[32] See Horberg et al. (2011), Prinz (2007: 68-86).
[33] See Prinz & Nichols (2010).
[34] Horberg et al. (2009).

involving emotion induction, and then move on to consider evidence gained from studies of patients with neurological abnormalities (*lesion studies*).

3.2.1 Evidence from emotion induction

To determine the causal relation between emotion and moral judgment, one natural procedure is to experimentally manipulate the former in order to determine its effect on the latter. Numerous studies of this kind show that emotion induction can modulate the intensity of our moral verdicts. Subjects induced to feel pangs of disgust in response to neutral words ('take', 'often') via posthypnotic suggestion were found by Thalia Wheatley and Jonathan Haidt (2005) to rate moral violations as worse when reading descriptions containing target words. Similar results have been obtained using a variety of disgust-inducing methods, including a canister of 'fart spray'.[35] Results of this kind are not restricted to the emotion of disgust. Jennifer Lerner and colleagues (1998) found that inducing anger intensified the degree of punishment subjects thought appropriate to transgressions described in fictional tort cases. Similar results are reported by Angelika Seidel and Jesse Prinz (2013).

More startling are cases in which emotional manipulation leads people to switch their overall moral verdict from 'permissible' to 'impermissible' or *vice versa*. In a second experiment conducted by Wheatley and Haidt (2005), subjects hypnotized to feel disgust at the words 'take' or 'often' were given a vignette describing Dan, a student council representative who 'often picks' or 'tries to take' discussion-topics that will appeal to students and faculty. When disgust-associated words were present in the vignette, some subjects expressed moral disapproval of Dan's actions. In a more recent study by Joshua Rottman and Deborah Keleman (2012), 7-year old children were shown pictures of anthropomorphic aliens performing a variety of unfamiliar actions,

[35] Schnall et al. (2008).

such as painting their faces white or filling a forest with cotton balls. The children were asked to say whether the described actions were 'wrong' or 'OK'. Some children were subject to disgust-induction via verbal and olfactory cues ('fart spray' again) and were significantly more likely to judge these apparently innocuous behaviours as morally wrong.

A person's affective state can also determine whether they regard actions as permissible or impermissible in the context of hypothetical moral dilemmas of the sort popularized by Philippa Foot (1967) and Judith Jarvis Thomson (1985), now staples of research on moral cognition. In the basic *Trolley Case*, a runaway trolley is on course to kill five people who cannot escape from a track up ahead; the trolley can be diverted onto a side-track, where it will kill only one person. In the *Footbridge Case*, a trolley is again on course to kill five and can only be stopped by pushing a large man off a footbridge into the path of the trolley, causing his death. Most people regard killing one to save five as permissible in the *Trolley Case* and impermissible in the *Footbridge Case*, though most people are also unable to provide an explanation for why there should be a moral difference.[36] It was hypothesized by Josh Greene and colleagues (2001) that people's differing verdicts are due to differences in the emotional responses elicited by these dilemmas: since harm must be caused in a manner that is 'up close and personal' in the *Footbridge Case*, the thought of killing one to save five here triggers a prepotent emotional response, leading subjects to condemn the action; the impersonal character of the original *Trolley Case* triggers no similar response.[37] Greene *et al.* obtained neuroimaging data consistent with this hypothesis.[38] Evidence

[36] Hauser et al. (2007).

[37] Greene (2008) has argued, in addition, that characteristically deontological moral judgments are driven by prepotent emotions, whereas characteristically utilitarian judgments rely on more dispassionate cognition. However, as noted by Kahane et al. (2011) the evidence amassed in support of this hypothesis is subject to a confound: once one controls for (counter)intuitiveness, there is no discernible difference in the neural signature associated with deontological and utilitarian moral judgments.

[38] Greene et al. (2001) cite reaction-time data in an attempt to demonstrate causality, but this data has

of causality emerges from an emotion induction study subsequently conducted by Piercarlo Valdesolo and Daniel DeSteno (2006). Subjects induced to experience a good mood by watching a comedy clip were found to be significantly more likely to view pushing the large man to his death as permissible in the *Footbridge Case*. No significant difference was observed regarding people's verdicts about the *Trolley Case*, suggesting that affect is uniquely important in determining people's moral judgment in the former.

3.2.2 Evidence from lesion studies

Studies of patients with neurological abnormalities can provide additional insight into the psychological mechanisms operative in the causation of certain behaviours. A range of studies involving subjects with damage to emotion-related brain areas provide further support for *Sentimentalism*.

Lesion studies confirm Greene's interpretation of subjects' differing reactions to the *Trolley Case* and *Footbridge Case*. Mario Mendez and colleagues (2005) put the *Trolley Case* and *Footbridge Case* to individuals with frontotemporal dementia, a degenerative neurological disorder involving deficits in affective processing. Unlike controls, patients with frontotemporal dementia tend to regard killing one to save five as permissible in both the *Trolley Case* and *Footbridge Case*. Similar results were obtained by Elisa Ciaramelli and colleagues (2007) and Michael Koenigs and colleagues (2007) using patients with damage to the ventromedial prefrontal cortex (VMPFC), a brain region associated with the integration of emotion in decision-making. The emotional 'alarm bells' ordinarily triggered when subjects contemplate pushing the large man to his death have apparently been silenced in frontal patients.

Patients with brain abnormalities also provide insight into the role played by

---

been found to be flawed by McGuire et al. (2009), as Greene (2009) concedes.

moral emotions in normal moral development. For patients who suffer VMPFC lesions in adulthood, the capacity for moral reasoning appears intact – even if certain hypothetical dilemmas can elicit nonstandard responses. By contrast, Steven Anderson and colleagues (1999) found that patients who suffer lesions during infancy exhibit abnormalities in moral reasoning, scoring at the 'pre-conventional' level in Lawrence Kohlberg's (1969) famous stage-sequence, tending to view conformity to moral rules as merely a strategy for avoiding punishment.

Additional evidence for the importance of emotions in normal moral development comes from studies of patients with psychopathy. Impairments in affective processing appear to be a central cause of psychopathic symptoms, with abnormalities of the amygdala and associated regions in the paralimbic system as a plausible source.[39] Psychopaths typically show little inhibition regarding immoral behaviour, failing to exhibit guilt or remorse and showing little empathy for their victims. As measured by skin-conductance, they show diminished arousal in response to images of human suffering.[40]

A range of studies conducted by Elliott Turiel and colleagues find that normally developing children across diverse cultures are able to distinguish between moral and conventional norms at ages as young as 3. Moral wrongs are rated as more serious, authority-independent, and generalizable, with conventional wrongs deemed less serious, dependent on established rules, and local to socio-cultural context.[41] The normal development of this capacity in human children appears to depend on the engagement of affect during social learning, as caretakers rely on empathy-induction when disciplining children's moral transgressions, adverting to the harm of the victim as a reason not to engage in proscribed behaviours.[42] This view is confirmed by the

[39] Blair et al. (2005); Kiehl (2008).
[40] Blair (1999); Blair et al. (1997).
[41] For overviews see Nucci (2001), Turiel (1983, 2006).
[42] Nucci & Turiel (1978); Smetana (1984, 1989).

finding that psychopaths, who are unmoved by the pain of others, do not follow the normal developmental trajectory. Using the same interview paradigm as Turiel *et al.*, incarcerated psychopaths were found by James Blair (1995) to exhibit no significant tendency to distinguish between moral and conventional wrongs in terms of their seriousness and authority-contingency; they were also significantly less likely to explain the wrongness of moral transgressions in term of harms to victims. Blair (1997) obtained similar results with children exhibiting psychopathic tendencies.[43] The capacity for sympathetic distress thus appears integral to normal moral development.

3.3 The origin of the moral emotions

We've now seen a range of evidence to suggest that our moral outlook is grounded to a large extent in our emotional responses. The aim of this section is to consider the origin of the moral emotions. Although some anthropologists regard all emotions as culturally-variable social constructions,[44] a tradition of research stretching back to Darwin (1872/2009) classifies a range of emotions as species-typical, biologically-based, and characterised by homologies in other animals.[45] The goal of this section is to provide support for a view of the moral emotions congruent with this research-tradition, with especial emphasis on homologies in closely related primate species.

The evidence I cite is designed to support the view of Frans de Waal and Jessica Flack (2000) that 'building blocks' of morality are observed in other primates and that human morality is an outgrowth of the social instincts that we share with

[43] In a recent study by Aharoni et al. (2012), psychopaths were asked to classify moral and conventional transgressions using a binary forced-choice format and performed better than chance. In light of Blair's studies, a natural interpretation of this data would suggest that psychopaths are able to mimic norm classifications without properly understanding their basis, much like a person blind from birth might know that strawberries are red and the sky is blue. In any case, Aharoni et al. find that higher scores on the affective component of psychopathy counterpredict accurate classification of norms.

[44] *E.g.*, Shweder (1994).

[45] Ekman (1972, 1992); Ekman & Friesen (1971); Griffiths (1997); Izard (1971); Levenson (1994, 2003); Matsumoto et al. (2008); Panksepp (1998); Tooby & Cosmides (1990). For critical discussion see Barrett (2006) and Prinz (2004). Mallon & Stich (1990) argue that disagreements between social constructionists and evolutionary theorists are primarily semantic.

other animals. Morality is thus not a radical break with our evolutionary past: human morality is a clear case of descent with modification. A similar view is presented by Jonathan Haidt under his *Moral Foundations Theory*.[46] According to Haidt, five domains of social action are consistently moralized across human societies: *Welfare/Harm*; *Fairness/Reciprocity*; *In-group Loyalty*; *Deference/Respect*; and *Purity/Sanctity*. Haidt regards each moral domain as grounded in a characteristic emotion or set of emotions: *e.g.*, sympathetic concern for *Welfare/Harm*; anger and gratitude for *Fairness/Reciprocity*; disgust for *Purity/Sanctity*. Haidt identifies precursors for four of these five domains in the emotional dispositions of other primates. The exception is the *Purity/Sanctity* domain. As disgust appears to be a uniquely human emotion,[47] moral concerns related to purity are supposed to be without antecedent in other animals. As I'll discuss in 3.3.4, this is not entirely correct: precursors for the moralization of incest are present in chimpanzees.

3.3.1 Harm and sympathetic concern

It appears that all cultures have harm norms of some sort.[48] These norms are otherwise highly variable along multiple dimensions, including the class of persons not to be harmed and the harms against which they are protected.[49] Harm norms exhibit a characteristic pattern of 'unity in diversity' shared by other moral domains: across different cultures certain high-level themes are constant, but specific rules differ markedly. While inconsistent with other forms of *Content Nativism*, this 'thematic clustering' is entirely consistent with the *Innate Biases Model*. Recall that the *Innate Biases Model* doesn't posit any determinate moral principles as innate, but points to facets of innate mental structure as biases that render moral norms relating to certain

[46] Haidt (2012); Haidt & Joseph (2004, 2007).
[47] See Kelly (2011), Rozin et al. (2008).
[48] Brown (1991); Haidt & Joseph (2004, 2007).
[49] See Prinz (2008).

issues more likely to arise and spread.

The moral emotion associated with harm is *sympathy* or *sympathetic concern.*[50] As evinced by Blair's studies of psychopaths, impairment of the capacity for sympathetic concern leads to abnormal moral development. Early development of sympathetic concern is consistent with the hypothesis of innateness. Several research groups have reported evidence of aversive responses to distress-cues in others that are present at birth,[51] with evidence of efforts to soothe distress in others for infants as young as 6 months[52] and consistent evidence of such behaviour for children aged more than 1 year.[53]

Evidence of sympathetic concern in *Pan* species is supported by multiple lines of evidence. Striking anecdotes are plentiful. Jane Goodall (1990: 213) notes that chimpanzees cannot swim, but make heroic efforts to save drowning conspecifics: in Gombe, one adult male died in an attempt to rescue an infant who had fallen into water.

More systematic data is available in records of consolation behaviour. If one chimpanzee attacks another, the aggressed party will sometimes seek comfort or reassurance from others. Goodall (1986: 361) reports that when a fourteen-year old male was attacked by a rival, he went to hold hands with his mother. Evidence of sympathetic concern is found in spontaneously initiated consolatory behaviour offered by unrelated bystanders. Based on hundreds of post-conflict observations, this consolatory behaviour is found to be moderated by exactly those variables we should expect if it is aimed at comforting victims.[54]

Evidence also indicates that chimpanzees intervene to prevent aggression against other members in their group: they experience a kind of *sympathetic anger*. In

[50] Nichols (2004); Nichols & Prinz (2010).
[51] Martin & Clark (1982); Sagi & Hoffman (1976); Simner (1971).
[52] Draghi-Lorenz et al. (2001).
[53] Zahn-Waxler et al. (1992).
[54] de Waal & Roosmalen (1979); de Waal & Aureli (1996).

the Arnhem Zoo, attacks by males against females are relatively common, but biting typically doesn't involve the sharp canines. If the canines are used, the high-pitched screams of attacked females can cause the entire colony to respond with an aggressive chorus of 'waa'-barks, sometimes followed by a group of females chasing off the offending male.[55] Among chimpanzees of the Budongo Forest in Uganda, third-party females are found to intervene in just under 30% of instances of male-on-female violence, using 'waa'-barks, threat gestures, and in some cases even violent retaliation.[56, 57]

3.3.2 Fairness and reciprocity

Cross-culturally, reciprocity represents another universal domain of moral concern in human beings.[58] Developmental evidence is consistent with the assumption of innateness, with evidence of a disposition to reciprocate in preschool children who have limited experience of complex, cooperative environments.[59] People are unusually adept at reasoning with conditionals related to matters of reciprocity, suggestive of an innate cognitive specialization for social exchange.[60] As regards the moral emotions, efforts to exploit social exchange are likely to elicit strong feelings of anger and resentment.[61]

Homologues of the social instincts which govern reciprocity in humans appear in chimpanzees. Captive chimpanzees are found to exercise dyadic reciprocity in

[55] de Waal (1991). See also de Waal (1982: 131, 1996: 91-92).
[56] Newton-Fisher (2006).
[57] Evidence of concern for the welfare of others is not unequivocal. Some studies find no evidence that chimpanzees care about the welfare of non-kin (Jensen et al. 2006; Silk et al. 2005; Vonk et al. 2008). However, more recent experiments using similar procedures have produced contrasting results (Horner et al. 2011; Melis et al. 2011). A key difference is likely to be that the latter allow the passive chimpanzees to communicate and solicit help from actors. For critical methodological discussion see Heyes (2012).
[58] Brown (1991); Fiske (1992).
[59] Levitt et al. (1985); Olson & Spelke (2008).
[60] Cosmides et al. (2010); Cosmides & Tooby (1992, 2008).
[61] Trivers (1971).

sharing food and grooming.[62] They punish 'stinginess': apes who do not reliably share food are more likely to meet with an aggressive response when attempting to obtain food from others. Chimpanzees also exercise reciprocity in relation to aggression: the tendency of one individual to support another in a fight varies with the tendency of the latter to support the former in turn.[63] This form of reciprocity is key to the coalitions formed by chimpanzees in negotiating status hierarchies.[64] Anecdote suggests that cheaters may be punished in this domain also. At Arnhem, a high-ranking female, Puist, helped a male friend, Luit, in chasing off a rival, Nikkie, but received no help from Luit when Nikkie turned on her: "Immediately after Nikkie had left the scene, Puist turned to Luit, barking furiously. She chased him across the enclosure and even pummelled him." (de Waal 1996: 97)

3.3.3 Equality in distribution

Human beings have moralized expectations about how goods are to be distributed. We can compare these expectations across cultures using a well-known paradigm from experimental economics, known as the *Ultimatum Game*. The first player in this game is able to divide a quantity of goods between herself and a second player; the other player can accept the division or reject it (in which case both get nothing). If both players are self-interested and rational (and this is common knowledge), the first player should allot the smallest possible amount to the second player and the second player should accept. Very few people behave in this way. In all cultures in which the game has been played, the second player tends to reject highly unequal offers, which tend not to be offered by the first player.[65] Players who reject positive offers report feeling anger at the greed of proposers. There is considerable variability across

[62] de Waal (1989).
[63] de Waal (1991, 1996: 156-157).
[64] See Mitani (2006), Watts (2002).
[65] See Camerer (2003).

cultures in the extent to which equal offers are proposed and unequal offers rejected.[66] Norms relating to the distribution of goods thus appear to follow the pattern of 'thematic clustering': the distribution of goods is moralized across all human cultures, but norms for distribution are otherwise variable. Studies of young children provide evidence consistent with innate foundations, showing that even infants expect goods to be distributed on an egalitarian basis.[67]

Adapting an experimental paradigm previously tested on macaques, Sarah Brosnan and colleagues (2005) were able to show that chimpanzees protest at the receipt of unequal rewards. Subjects had previously been taught to exchange tokens for food rewards. Chimpanzees were tested in pairs to determine whether unequal rewards would induce protests and refusals. Disadvantaged chimpanzees showed clear signs of reacting aversively when their partner received a more desirable food item. A control condition established that it was not merely the salience of a more valuable reward driving these protests: disadvantaged subjects were protesting at the inequality itself.[68]

3.3.4 Beyond harm and fairness

We've now seen evidence of innate building blocks relating to norms of harm and justice in closely related primate species: chimpanzees are sensitive to the welfare of others, they experience anger when they are cheated in social exchange, and they protest when they are on the losing end of an unfair distribution. Haidt reminds us that we shouldn't ignore other potential facets of the moral domain. Issues relating to

[66] Henrich et al. (2005); Henrich et al. (2010).
[67] Geraci & Surian (2011); Sloane et al. (2012). See also LoBue et al. (2009) and Olson & Spelke (2008) for evidence of egalitarian concern in pre-school children.
[68] Efforts have also been made to adapt the *Ultimatum Game* to be played by chimpanzees, but these yield conflicting results and are subject to fierce methodological controversy. See Henrich & Silk (2013), Jensen et al. (2007), Jensen et al. (2013), Proctor et al. (2013), Smith & Silberberg (2010).

welfare and justice just about exhaust the concerns of Western liberals,[69] who comprise the majority of academic researchers. Other groups emphasize a broader suite of moral concerns. As noted, Haidt posits three additional domains of social action as consistently moralized across human societies: *In-group Loyalty*; *Deference/Respect*; and *Purity/Sanctity*. Haidt believes that affective dispositions inherited from the last ancestor of *Homo* and *Pan* bias the evolution of norms in all domains except *Purity/Sanctity*. Thus, the moralization of *Deference/Respect* is thought to be rooted in social instincts relating to dominance and submission exhibited in the despotic hierarchies characteristic of chimpanzee societies. *In-group loyalty* is supposed to have its roots in forms of 'community concern' observed in chimpanzees. Chimpanzees appear to care about the general quality of social relationships throughout their community, as evinced by efforts by third-parties to mediate conflicts and bouts of group-wide celebration that occasionally follow successful reconciliation.[70]

As noted earlier, Haidt and Joseph (2007) suggest that 'building blocks' of the *Purity/Sanctity* domain are absent in other primates. Their reasoning appeals to the close link between purity and disgust, in combination with the fact that disgust appears to be a uniquely human emotion. However, affective dispositions shared with *Pan* species seem to bias moral judgments relating to at least one issue falling under Haidt's *Purity/Sanctity* domain: incest.[71]

Cross-culturally, norms against incest follow a familiar pattern. Norms which prohibit sexual contact among members of the immediate family are prevalent, but explicit prohibitions are not recorded in all societies.[72] Explicit permissions are

[69] See Graham et al. (2009).
[70] de Waal (1996: 163-208); de Waal & Flack (2000: 14-16).
[71] Kelly (2011) argues that although disgust is uniquely human, it results from the co-evolution of two independent affect systems which are present in other animals.
[72] Brown (1991).

recorded in some instances.[73] Many societies moralize sexual contact with more extended family, such as first cousins, but there is a high degree of cultural variability in this dimension and cousin marriage is common in many societies.[74]

Edward Westermarck (1891/1922) hypothesized that incest norms have their origin in adaptive aversion to sexual contact amongst members of the immediate family. He further hypothesized that because genetic relatedness is not immediately observable, childhood co-residence should serve as a cue for incest aversion. A psychological disposition to avoid sex with childhood co-residents is now called the *Westermarck Mechanism*. While controversial for some time, Westermarck's conjecture now enjoys substantial support.[75] Debra Lieberman and colleagues (2003, 2007) have shown that childhood co-residence is positively correlated with greater disgust at sexual contact with one's siblings and with greater moral disapproval of incest amongst third-parties. Incest avoidance is observed in the vast majority of animal species,[76] including chimpanzees and bonobos.[77] Because the psychological mechanisms which underwrite incest-avoidance appear to be shared with *Pan* species, it seems mistaken to suppose that moral concerns falling under the *Purity* domain are entirely without antecedent in other primates.

3.4 Summary

In this section, I've been building a case for the view of human morality as influenced by a suite of innate affective biases inherited from the last common ancestor of *Homo* and *Pan*. To support this view, I began by marshalling evidence that people's emotional states influence their moral outlook via a multiplicity of pathways. I then sought to

[73] Prinz (2008).
[74] Bittles (1990).
[75] Lieberman (2008).
[76] Parker (1976).
[77] de Waal (2001: 342-343).

demonstrate that chimpanzees exhibit affective dispositions corresponding to widely moralized social behaviours, pointing also to the existence of similar innate dispositions in human children. In light of the close relatedness between *Homo* and *Pan,* these similarities are indicative of homology: these traits most likely derive from precursors present in our last common ancestor. Given the evidence for *Sentimentalism* already considered, the correspondence I've noted between these dispositions and certain cross-cultural domains of concern is plausibly explained by the tendency of the former to bias the latter. Hence, we have reason to suppose that human morality reflects a form of social life and a set of congruent emotional responses that we share with closely related African apes.

4. Morality as a group-level functional trait

4.1 Introduction

In this section, I'm going to present evidence in support of the second proposal that I set out in my introduction: that moral norms characteristic of human societies are cultural systems of social control adapted by group selection to promote cooperation and suppress competition.

In my view, the evolution of morality is tightly interwoven with that of large-scale cooperation. Human beings cooperate extensively with non-kin,[78] whereas similar behaviour in chimpanzees is limited.[79] As Jonathan Haidt (2012) puts it, we are 90% chimp and 10% bee. Our focus in this section will be on that 10%. As we'll see, converging lines of evidence support the view that the full extent of human cooperation cannot be explained without appeal to some form of group selection. I will support the view that the evolution of morality is tightly interwoven with this

[78] Bowles & Gintis (2011); Richerson & Boyd (2005)
[79] Muller & Mitani (2005).

process of group competition. The underlying idea here is little different from that expressed by Darwin in *The Descent of Man* (1879/2004):

> It must not be forgotten that although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe, yet that an advancement in the standard of morality and an increase in the number of well-endowed men will certainly give an immense advantage to one tribe over another. … At all times throughout the world tribes have supplanted other tribes, and as morality is one important element in their success, the standard of morality and the number of well-endowed men will thus everywhere tend to rise and increase. (157-158)

This approach has recently been given new theoretical and empirical foundations,[80] on which I draw extensively. I'm going to start by outlining the importance of group competition in explaining human cooperation. In section 4.2, I argue that certain forms of prosocial behaviour cannot be explained without appeal to group selection. Section 4.3 offers independent evidence that should lead us to expect that group competition has been integral to the evolution of human cooperation. Finally, in section 4.4, I outline the role played by moral systems in this process. I rely on two 'case studies': the role of moral norms in coordinating the punishment of free riders, and the role of egalitarian norms in reducing differences in fitness within groups.

4.2 A cooperative species

We might expect natural selection to favour something approaching ruthless selfishness. An organism that benefits another at a cost to itself appears at a distinct disadvantage in the struggle for existence. Even costless behaviours that increase the fitness of others are liable to be selected against, because natural selection is a matter of relative – not absolute - fitness. Nonetheless, nature is not unrelentingly cut-throat.

[80] See Boehm (1999), Bowles & Gintis (2011), Henrich & Henrich (2007), Richerson & Boyd (2005), Sober & Wilson (1998).

Altruism is found even in unicellular organisms.[81]

Evolutionary theorists have devoted considerable energy to explaining the costly prosocial behaviours observed in nature. Prior to the 1960s, appeals to 'the good of the group' were routine in this respect, but theoretical work beginning in that decade allowed evolutionary biologists to identify a number of competing explanations. My aim here will be to argue that no explanation of large-scale cooperation among human beings is adequate unless some appeal is made to competition among groups. To make my case, I'm going to consider five popular mechanisms for explaining the evolution of cooperation that have emerged from the 1960s onward and outline why they are insufficient in this respect. The mechanisms I consider are *kin selection*, *direct reciprocity*, *indirect reciprocity*, *costly signalling*, and *punishment*.

I'm especially going to emphasize their shortcomings in the context of public goods provisions. Human beings reliably work together in large groups for the sake of a common good: we vote; we recycle; we defend our societies against external threats. Because these goods are public and cannot easily be restricted from those who don't contribute, they are vulnerable to free riding. Nonetheless, human beings reliably exhibit a disposition to cooperate in such contexts. As with people's expectations about distribution, we can examine behaviour of this kind using economic games. Participants in the *Public Goods Game* are given a private account containing some amount of money and then have the opportunity to deposit some of this money into a public pool; they are also allowed to keep everything they do not contribute to the pool. The experimenter multiplies the total amount in the common pool by some value (*e.g.*, 1.5) and the resultant amount is then distributed equally among the players. The pay-offs and number of participants are orchestrated so that the dominant strategy for

[81] See Hudson et al. (2002) on amoebae.

an entirely self-interested player is to contribute nothing to the public pool. The behaviour expected of self-interested players is rarely observed in human beings.[82]

Whereas the five mechanisms we'll consider are well-suited to explain a variety of cooperative behaviours, they are ill-suited to explaining cooperation among unrelated individuals in the provision of public goods. As we'll see, kin selection and direct reciprocity are essentially non-starters in this respect. The remaining mechanisms share a similar flaw: we cannot explain their consistent association with public goods provisions unless we invoke some mechanism for equilibriums selection. Group selection is the most natural solution to this problem.

4.2.1 Kin selection

We begin with kin selection. Because genes exist in multiple copies located across numerous individuals, they can promote their own replication by coding for traits that raise the fitness of other organisms carrying the same genes. Actions that come at some cost to the fitness of individuals can thus be favoured by natural selection if they are directed at close relatives. Following, Maynard Smith (1964), we refer to this process as 'kin selection'.

It's clear that kin selection explains a range of altruistic behaviours observed in our species. People are willing to make considerable sacrifices for the sake of close relatives. For example, in the United States, 86% of donated kidneys come from close kin.[83] However, kin selection is otherwise of limited explanatory value when it comes to cooperation among humans, because we cooperate extensively with non-kin. Even in the smallest of hunter-gatherer bands, the average degree of relatedness is low and many individuals are entirely unrelated.[84]

[82] Camerer (2003); Henrich et al. (2005).
[83] Henrich & Henrich (2007).
[84] Chudek et al. (2013).

4.2.2 Direct reciprocity

The theory of direct reciprocity can explain cooperation among unrelated individuals in dyadic contexts, provided there is some chance of repeated interaction. The theory was originally worked out by Robert Trivers (1971) with reference to the *Prisoner's Dilemma*. The benefits that could be gained from cooperation in the *Prisoner's Dilemma* are undercut by the vulnerability of co-operators to defectors: since the worst possible outcome is to cooperate while the other player defects, defection is the dominant strategy, although payoffs are higher if both cooperate than if both defect. This problem can be overcome if the game is repeated over several rounds, yielding an *Iterated Prisoner's Dilemma*. In this setting, the cooperative inclinations of others can be assessed on the basis of past interactions, allowing individuals to cooperate with others who are similarly inclined and withhold cooperation from defectors. Strategies for conditional cooperation can then evolve due to the higher pay-offs gained from reciprocal cooperation.

One of the most straightforward conditional strategies designed to take advantage of this form of reciprocal cooperation is *Tit-for-Tat*, which became a focus for early work in this area.[85] Individuals playing this strategy cooperate on the first round of the *Iterated Prisoner's Dilemma*, cooperate on the next round if they received cooperation on the previous round, and otherwise defect. Subsequent work has examined a range of alternate strategies that outcompete *Tit-for-Tat* under realistic conditions.[86]

While well-suited to explaining cooperation in dyadic contexts, direct reciprocity proves ill-equipped to explain cooperation in public goods settings, except

[85] Axelrod (1980a, 1980b); Axelrod & Hamilton (1981).
[86] Nowak & Sigmund (1992, 1993, 1994); Roberts & Sheratt (1998). For overviews see Henrich & Henrich (2007) and Trivers (2006).

in very small groups.[87] In dyadic contexts, strategies of reciprocal cooperation succeed because it's possible to direct cooperative behaviour toward other co-operators and withhold it from defectors. The same is not possible when dealing with public goods: the only means of withholding cooperation from defectors is to withhold cooperation from the group as a whole, and thus from other co-operators.

4.2.3 Indirect reciprocity

Under indirect reciprocity, cooperative behaviour evolves because it is reciprocated by third-parties, rather than by its immediate recipients.[88] A system of indirect reciprocity can thus provide an incentive for people to behave cooperatively even when there is no opportunity for repeated interaction between any two individuals. As in the case of direct reciprocity, multiple strategies have been designed to take advantage of this mechanism, and there remains considerable debate as to which are more likely to evolve under realistic conditions.[89]

Karthik Panchanathan and Robert Boyd (2004) have shown that cooperation in public goods settings can be stabilized under certain conditions if linked to a system of indirect reciprocity. However, this result is subject to a notable limitation: to the extent that indirect reciprocity is able to sustain public goods contributions, this has nothing to do with the benefits conferred by contributing to the public pool. An analogous system of indirect reciprocity could just as well stabilize any other similarly costly behaviour. Thus, any attempt to explain human cooperation in public goods settings by reference to indirect reciprocity is incomplete unless buttressed by some explanation for why public goods contributions in particular should be stabilized via this mechanism.

---

[87] Bowles & Gintis (2011); Richerson & Boyd (1998).
[88] Alexander (1987).
[89] Leimar & Hammerstein (2001); Milinski et al. (2001); Nowak & Sigmund (1998); Panchanathan & Boyd (2003); Sugden (1986).

4.2.4 Costly signalling

A similar problem applies when it comes to explaining cooperation by means of costly signalling. Organisms often need to advertise characteristics that are not directly observable, such as their quality as a mate or the extent to which they'll fight for valuable resources. For a signal to be useful, it must be hard to fake. One solution to this problem is Amotz Zahavi's (1975) *Handicap Principle*: the signal is kept honest because it is too costly to produce unless one has the relevant hidden characteristic. While initially controversial, the *Handicap Principle* is now thought to explain a number of characteristics that otherwise appear debilitating, especially in the context of sexual selection.[90] The elaborate train of the peacock is the paradigm example.

The theory of costly signalling has also been proposed as a potential explanation for prosocial behaviour in human beings and other animals: costly helping behaviours are supposed to have evolved because they act as honest signals of desirable unobservable traits, such as mate quality.[91] Consistent with this assumption, male players contribute more in the *Public Goods Game* if they are in the presence of an attractive female.[92]

Herbert Gintis and colleagues (2001) have shown that under certain plausible conditions, cooperation in public goods settings can be stabilized as a form of costly signalling and can spread from a condition of initial rarity. However, as with indirect reciprocity, the ability of signalling to stabilize cooperation is independent of the group benefits conferred by public goods contributions: what matters for the success of the signal is its costliness to the actor, not its benefits to third-parties. Thus, to explain large-scale cooperation by reference to costly signalling some supplementary

[90] See Maynard Smith & Harper (2003).
[91] Miller (2000); Van Vugt & Iredale (2013); Zahavi (1975, 1995).
[92] Van Vugt & Iredale (2013).

explanation must also be provided for why signals that involve public goods contributions should reliably emerge.

4.2.5 Costly punishment

The final mechanism for explaining the evolution of cooperation that we'll consider is the use of punishment. By invoking punishment, we're already straying onto the issue of the benefits conferred by moral systems, as moral norms are the typical proximate mechanism underlying punishment of free riders in human societies. I'm going to shelve a more focused discussion of the role of morality in the evolution of cooperation until section 4.4. For now, I want us to concentrate on evaluating punishment as a solution to the problem of free riding in public goods contexts.

In the absence of the option to punish those who free-ride, contributions in the *Public Goods Game* actually tend to decline if the game is repeated over several rounds.[93] It appears that initially cooperative players decide to withdraw cooperation when they detect the presence of free-riders in the group. When it becomes available, subjects across different cultures reliably favour the punishment option, and the availability of punishment substantially increases contributions to the common pool.[94] While it is thus very natural to give punishment a prominent role in explaining cooperation in public goods settings, it is less obvious how to develop this story in a plausible manner. Where punishment serves to increase the availability of public goods contributions, it becomes a higher-order public good, subject to its own cooperative dilemma. As a means of explaining public goods contributions, costly punishment might therefore appear to be a nonstarter, merely reproducing the explanatory problem one level up.

Even if the use of punishment is never costless in expectation, punishment will

[93] Fehr and Gächter (2002); Page et al. (2005).
[94] Fehr and Gächter (2002); Herrmann et al. (2008); Ostrom et al. (1992); Rockenbach & Milinski (2006).

typically be *less* costly than the primary behaviour it's designed to reinforce.[95] As a form of public goods provision, punishment will therefore often be cheaper than corresponding first-order contributions. Furthermore, to the extent that punishment succeeds in inducing individuals to engage in cooperative behaviour, the disposition to punish becomes less costly overall as the target behaviour is rarely seen. For similar reasons, a disposition to punish becomes less costly as the frequency of punishers increases. Because the advantage to free riding is thus relatively minor when it comes to punishment, it may be easier to explain the evolution of large-scale cooperation backed by costly punishment via mechanisms like those already discussed in previous subsections.[96]

In addition, Robert Boyd and Peter Richerson (1992) have shown that second-order punishment can stabilize public goods contributions under certain conditions. When the costs of being punished grow large enough, the use of higher-order punishment can create an evolutionarily stable strategy in which individuals cooperate in public goods settings, punish defectors, and *also* punish those who do not punish. However, the ability of higher-order punishment to stabilize cooperation is again independent of the benefits conferred by public goods cooperation: a similar system of punishment can in principle stabilize any behaviour, provided the costs of punishment are high enough. We thus face the same problem of explaining the consistent use of punishment to stabilize public goods cooperation. In addition, evidence that human beings actually rely on higher-order punishment does not appear forthcoming.[97]

4.2.6 A role for weak group selection

We've seen that indirect reciprocity, signalling, and punishment can in principle

[95] Sober & Wilson (1998).
[96] There is some evidence that punishment acts as a form of costly signalling: Barclay (2006); Kurzban et al. (2007).
[97] See Kiyonari & Barclay (2008), Wiessner (2005).

stabilize cooperation in public goods settings. However, in each case a question arises as to why this behaviour in particular should be stabilized via this mechanism. There is a straightforward solution in each case: weak group selection. As I've noted, where several behaviours can be stabilized by intragroup processes, selection across groups may favour the evolution of some ahead of others. When it comes to explaining why we find the use of costly punishment consistently recruited in the service of increasing public goods provisions, the group pay-offs provide a natural answer: members of groups in which punishment is used in this way are fitter. Absent some other plausible mechanism for equilibrium selection, explanations of large-scale cooperation in human beings will thus fall short unless supplemented by the assumption that cooperation has evolved in the context of group competition.

4.2.7 A role for strong group selection?

An additional shortcoming applies to all of the models discussed above. In economic games, people reliably confer costly benefits on others to whom they are unrelated, even when interactions are unrepeatable and anonymous. The cost cannot be repaid by the recipient at a later point, nor can it enhance the actor's reputation, signal her fitness, or allow her to evade punishment. *Prima facie*, there is no way to explain these behaviours using any of the models we've considered so far - except by assuming that they are 'misfires' produced by behavioural strategies adapted to environments in which interactions of this kind were of negligible importance. For example, if the ancestral environment during the Late Pleistocene and Early Holocene was one in which all interactions were highly likely to be repeated, we might have evolved to be insensitive to the possibility of one-shot encounters. Similarly, if social behaviour was always observable by third parties, we might have evolved behavioural strategies that operate on the default assumption that one's reputation is at stake. Behaviour in

economic games would then be an example of *adaptive lag*, the phenomenon whereby an organism is disadvantaged due to a mismatch between its current environment and the environment in which its species evolved. The view that behaviour in economic games is due to adaptive lag is accepted by many as an explanation for behaviour observed in economic games,[98] but has attracted considerable skepticism.[99]

Although it is difficult to resolve this issue with much certainty, I believe that skepticism is warranted. There is ample evidence for the occurrence of one-shot encounters among hunter-gatherers, including reliable encounters with strangers during migrations.[100] In addition, evidence indicates that ancestral bands will have left considerable room to engage in proscribed behaviours beyond the prying eyes of others. Infidelity is a case in point. Extra-pair copulations are among the most divisive transgressions in hunter-gatherer societies, being the primary cause of intragroup homicide.[101] There is a very strong incentive to monitor and suppress such behaviour. Nonetheless, infidelity occurs reliably and evidence suggests that opportunities for extra-pair mating have exerted significant selection pressure on human psychology and physiology. Human females exhibit behavioural strategies that appear designed to take account of extra-pair copulations to improve the genetic quality of their offspring.[102] We also find compensating adaptations in human males, including the size of the testes and the shape of the penis.[103] Opportunities for proscribed behaviours to occur in secret have therefore occurred with sufficient reliability to have exerted a recognisable selection pressure on human psychology and physiology.

In sum, ancestral societies were not as close-knit or as closely observed as the appeal to adaptive lag would suggest. We should thus be at least somewhat doubtful

[98] See Cosmides & Tooby (1992), Dawkins (2006), Trivers (2006).
[99] Bowles & Gintis (2011); Fehr & Henrich (2003); Chudek et al. (2013); Richerson & Boyd (2005).
[100] Fehr & Henrich (2003).
[101] Boehm (2000).
[102] Buss (2000); Gangestad (2006).
[103] Shackelford & Goetz (2007).

that the full extent of human prosocial behaviour can be explained by any of the models discussed in this subsection, even if we assume these to be supplemented by a process of weak group selection. To fully account for human cooperation, some degree of strong group selection may be required. As I'll discuss in the next subsection, we have independent reason to suppose that the conditions for the operation of strong group selection were present during the Late Pleistocene and Early Holocene.

4.3 Conditions for group selection

The previous subsection offered reasons to suppose that group competition has played a significant role in the evolution of human cooperation. Here, I point to two conditions which have potentiated group competition within human evolutionary history: the emergence of large-scale cultural differences governed by conformity biases and the existence of high rates of intergroup violence.

4.3.1 The role of culture

Natural selection requires phenotypic variation in fitness-relevant traits: where such differences are great, selection is strong; where they're slight, selection is weak. One of the reasons that biologists have been prone to dismiss group selection as an evolutionary force is that stable genetic differences among animal groups are typically slight and easily offset by migration and exogamy.[104]

Increasing interest in the role of culture in human evolution has been one of the key contributing factors to the rising fortunes of group selectionist explanations for human cooperation. This is in large part due to the ability of culture to amplify group differences. Genetically similar groups inhabiting nearly identical habitats can exhibit significant phenotypic differences due to culture, which allow some to expand

[104] See Maynard Smith (1964), Williams (1966).

at the expense of others. Not only does the role of culture in human societies increase differences between groups, it also serves to flatten variation within groups, thus diminishing the force of intragroup selection. Cultural transmission is subject to a number of biases, including *context biases*, which increase the likelihood of cultural transmission due to the identity of those already exhibiting the trait. One such bias is the *conformity bias*, which leads individuals to copy the behaviour of the majority.[105] The strength of this bias is illustrated in Solomon Asch's (1952) famous conformity experiments, in which subjects readily endorse obviously incorrect perceptual judgments about the relative length of two lines in order to go along with the majority.[106] The existence of conformity bias leads members of the same group to behave similarly: outliers don't remain outliers for long as common traits become more common and rare traits rarer still. Conformity bias also renders cultural traits stable against the force of migration, with new arrivals adopting the behaviour of the majority.

4.3.2 Intergroup violence

At the time that Williams, Maynard-Smith, and Dawkins were ostensibly burying group selection as a plausible mechanism for the evolution of altruism, it was widely believed that hunter-gatherers are generally peaceful.[107] Although this issue remains controversial, evidence increasingly suggests that a romantic view of pre-state societies is unwarranted. Pre-state societies appear to be characterised by high frequencies of deaths due to intergroup violence, at levels that outdo even the bloody 20th century.

Violence among small-scale societies typically takes the form of intermittent,

[105] Boyd & Richerson (1985).
[106] For cross-cultural replications see Bond & Smith (1996).
[107] See Ember (1978).

small-scale raids with limited casualties. However, as casualties pile up following successive raids, groups eventually break up, cede their territory, and become extinct. Sam Bowles (2009) finds that both archaeological and ethnographic sources indicate a mean rate of adult death due to warfare of 14%. By comparison, it is estimated that the share of deaths attributable to wars among industrialized societies during the 20th century is just 0.7%.[108] Such levels of intergroup violence are likely to have made the force of intergroup selection considerable. This expectation is confirmed by a model constructed by Bowles to simulate warfare in the Palaeolithic, aimed to evaluate Darwin's suggestion that altruism could evolve by affording groups an advantage in intergroup contests. Bowles combined data for the rate of intergroup killings with estimates of the degree of genetic variation among ethno-linguistic groups drawn from extant hunter-gatherer populations to construct a model of group competition that might reasonably approximate conditions during the Late Pleistocene and Early Holocene. He found that the rate of intergroup warfare was sufficient to support the evolution of altruism under a variety of reasonable parameter-settings. Note that this model delimits phenotypic variation to genetic differences, thus (deliberately) understating the extent of intergroup differences.

4.3.3 Supporting Evidence

Contemporary social psychology provides additional corroborating evidence that human cooperation has evolved in the context of group competition.

Human beings are strongly inclined to carve up the social world into competing tribes, suggesting a heightened sensitivity to group differences. The ease with which we fall into an *us-vs.-them* mentality is illustrated by experiments employing the *minimal group paradigm* pioneered by Henri Tajfel (1970). These

[108] Pinker (2011).

experiments show that it is possible to trigger ingroup favouritism based on the most arbitrary of classifications: for example, Tajfel and colleagues were able to induce an ingroup bias by assigning subjects to two different groups based on whether they tended to over- or underestimate the number of dots flashed on a screen. The readiness with which people divide the world into competing coalitions and develop ingroup biases has been taken to suggest that we have an innate cognitive specialization for negotiating a landscape marked by group competition.[109] Consistent with this hypothesis, even infants show sensitivity to ethnolinguistic cues, preferring to look at faces belonging to their own race and favouring foods offered by speakers of their native language.[110]

As noted by Bowles and Gintis (2011), further supporting evidence emerges from the discovery that prosocial behaviour and ethnocentric tendencies rely on shared neural mechanisms: the neuropeptide oxytocin reliably increases both prosocial motivation and ingroup bias. The tendency of oxytocin to increase bonding, generosity, empathy, and trust has been widely publicised.[111] However, the so-called 'cuddle chemical' has a dark side. Carsten De Dreu and colleagues (2011) found that nasally administered oxytocin increased in-group favouritism and out-group derogation in Dutch subjects. For example, subjects who received oxytocin showed a stronger tendency to implicitly associate Dutch names with positive words, and a stronger implicit association between negative words and Arab names.

4.4 The role of morality in group competition

We've now seen a broad range of evidence to support the view that group selection has played an important role in the evolution of large-scale cooperation in our species.

[109] See Gil-White (1999, 2001), Haidt & Kesebir (2010), Richerson & Boyd (2005), Spelke & Kinzler (2007).
[110] Spelke & Kinzler (2007).
[111] See Macdonald & Macdonald (2010) for an overview.

What role have moral norms played in all this?

The general answer is that morality has functioned as a system of social control: a means of ordering group life so as to suppress competition and increase cooperation. By altering the balance of costs and benefits associated with prosocial and antisocial behaviours, moral norms have contributed to the constellation of factors favouring the evolution of group-beneficial behaviours. If moral norms are enforced with sufficient efficiency, the relevant pay-offs may be altered in such a way that an otherwise altruistic behaviour becomes optimal from the perspective of maximizing individual fitness. Even if norm-enforcement is not quite so effective and the target behaviour remains altruistic in some respect, the effect of moralization is to reduce the disadvantage of altruists to a point at which the behaviour in question becomes easier to evolve under strong group selection.

The position advanced here is, I believe, highly intuitive. As the eye appears designed for sight, the moral systems characteristic of human societies appear designed to mould groups into well-functioning units.[112] In the following, I'll pick out two examples that serve to illustrate this general picture of morality as a group-level adaptation. The first fulfils a promise made earlier in this section, by elaborating on the role of moral norms in policing free riding. The second looks at egalitarian norms, arguing that they are a device for reducing fitness disparities within the group.

4.4.1 Policing free riders

We've already noted one key role played by moral norms in group competition: moral norms are the proximate mechanisms underlying the policing of free riders in public goods contexts. Cooperation in the provision of public goods is consistently moralized across human societies, with free riders subject to sanctions ranging from harsh words

[112] See Sober & Wilson (1998: 175-181).

to ostracism.[113] The practice of sharing meat among hunter-gatherers is a case in point.[114] Because gains from hunting are variable, sharing game throughout the group acts as a form of insurance. Everyone is better off under this practice of risk pooling, though anyone could do better if they were able to free-ride on the system. Consequently, meat sharing is actively encouraged through vigilant group pressure, and perceived stinginess readily invites moral criticism. In general, stinginess, greed, and laziness are among the most frequent topics for moral censure among hunter-gatherers.[115]

As we recall, the use of punishment appears essential to maintaining cooperation in public goods settings: without the option to punishment, would-be cooperators gradually opt out in order to withhold benefits from free riders. We may therefore expect selection to have favoured proximate mechanisms that promote the targeted punishment of free riders as a means of sustaining public goods contributions. Moral norms play just that role. As moral sanctions plausibly involve a form of costly policing, their use is subject to a second-order free riding problem, but we have already outlined a variety of means by which this problem could be solved.

In the context of policing free riding, moral norms have an especially important role to play in terms of coordinating responses. Although the use of punishment reliably increases overall contributions to the public pool in the *Public Goods Game*, there is a complication to this story that I've so far suppressed: the use of punishment is often found to *decrease* group payoffs overall.[116] The decline occurs because the costs incurred in the use of punishment outweigh the benefits of added public contributions. Not only do players who make high contributions to the public pool happily punish free riders, free riders also engage in retaliatory punishment of

[113] Sober & Wilson (1998); Wiessner (1996).
[114] Kaplan et al. (1984); Boehm (1999).
[115] Wiessner (2005).
[116] Dreber et al. (2008); Egas & Reidl (2008); Herrmann et al. (2008); Ostrom et al. (1992); Rochenbach & Milinski (2006).

high-contributing players. When the costs of punishing and being punished are weighed against the benefits of greater contributions to the public pool, the result is a net loss in payoffs.

The problem of declining group payoffs can be overcome if the punishment condition is modified so as to greater approximate the moralized social control characteristic of human communities. Effective punishment requires perceived legitimacy. Experiments where participants are allowed not only to punish each other but also to devise shared norms for behaviour in the *Public Goods Game* lead to greater cooperation *and* higher groups pay-offs.[117] The effective use of punishment thus depends on standards of public legitimacy without which we should expect group selection to counteract the evolution of punishment in public goods settings. In human societies, moral norms perform just this function.

4.4.2 Egalitarianism as fitness levelling

Another important respect in which moral norms have served to alter the costs and benefits of cooperative behaviours is through their tendency to enforce a form *fitness levelling*: reducing disparities in reproductive success that would otherwise emerge due to competition for status.[118]

Mechanisms to suppress fitness differences occur elsewhere in biology as a means of delimiting harmful competition. One well-studied case is the genome. We are used to thinking of genes located in different organisms as competing against one another. Competition among alleles located in the same individuals also occurs. In diploid heterozygotes, individuals carry two different alleles of the same gene, located on homologous chromosomes. The sex cells are haploid: they contain only one

[117] Ertan et al. (2009); Ostrom (1992).
[118] For this view see Alexander (1987), Boehm (1999), Bowles (2006), Frank (2003), Sober & Wilson (1998).

chromosome from each pair of homologues and thus contain only one variant of the gene. Intragenomic selection favours alleles that are able to ensure a greater representation among gametes. Unfortunately, *distorter alleles* of this kind are often costly to the fitness of individuals, and so to the reproduction of genome as a whole. Genomic selection has thus favoured compensating mechanisms designed to ensure *fair meiosis*, with alleles on homologous chromosomes having roughly equal chances of inclusion in gametes.[119] The only way for an allele to increase its fitness is then to work for the 'good of the group': increasing the likelihood that the genome as a whole is replicated.

In human beings, egalitarian norms appear to perform a similar function. As noted in 2.3, social life in many primate species is governed by a despotic hierarchy in which dominant males enjoy privileged access to resources and mating opportunities. Dominance is a key determinant of evolutionary fitness and political manoeuvring aimed at rising up the hierarchy is intense and ongoing.[120] Not all primates live in hierarchical societies: squirrel monkeys are a case in point.[121] Our own species is difficult to classify because we vary considerably according to social conditions. Post-Neolithic societies like our own are significantly stratified, though ideals of equality may remain prominent in moral discussions. Nomadic hunter-gatherers are egalitarian in principle *and* practice. Meat is shared between families on an egalitarian basis. Groups have no established leaders and rely on consensus-seeking.[122] Egalitarian social relations are actively maintained through sanctions and moral criticism.[123] Efforts by individuals to assert authority over the group by assuming a position of leadership meet with mockery and disobedience; aggressive and violent men who persist in seeking dominance are liable to meet with ostracism and even execution.

---

[119] See Leigh (1977) for further discussion.
[120] As explored in de Waal (1982).
[121] Boinski (1994).
[122] Mithen (1990); Knauft (1991).
[123] Boehm (1999, 2000).

The overall effect of the moral commitment to egalitarianism exhibited by hunter-gatherers is to reduce disparities in fitness within the group by eliminating the privileges associated with alpha-status and otherwise curtailing differences in reproductive success associated with differing ranks. By levelling fitness, egalitarian norms serve to mitigate the force of intragroup selection and favour the evolution of group beneficial behaviours. In fact, it is difficult to explain their prevalence otherwise: efforts to simulate group competition in ancestral environments indicate that the cost to groups of upholding redistributive norms should lead to their elimination under group selection *unless* the practice of fitness levelling serves to promote group-beneficial behaviours.[124] Egalitarian norms are thus plausibly viewed as a group-level adaptation.

4.6 Summary

This completes my case for the second hypothesis set out in my introduction. Human beings cooperative extensively outside the kinship circle and build complex societies to rival eusocial species like the honey bee and naked mole rat. The evolution of large-scale cooperation in human beings is difficult to explain without invoking group competition. Because people cooperate even in anonymous, unrepeatable contexts, there is some reason to suppose a role for strong group selection, and this view is further corroborated by evidence of favourable evolutionary conditions, including cultural differences that amplify variation across groups, high-rates of death from warfare, and the widespread practice of fitness-levelling egalitarianism. Within the context of group competition, social control underwritten by moral norms is likely to have offered considerable advantages by coordinating punishment and altering the balance of costs and benefits to favour group-beneficial prosocial behaviours. The

[124] Bowles (2006).

systems of moral norms characteristic of human societies are thus properly viewed as adaptations evolved under group selection.

5. Conclusion

The claim that morality is the product of evolution is essentially truistic: there is virtually nothing in biology which is not in some sense due to evolution. In this chapter, I've outlined and defended two specific proposals that attach evolutionary explanations to certain key elements of our moral psychology. Firstly, I have argued that our moral outlook is influenced by innate affective biases with recognisable homologues in closely related primate species, suggesting that our moral concerns reflect the particular conditions of hominine social life. Secondly, I have argued that the kinds of moral norms we observe across human societies are not merely by-products, but have actively been selected for in the context of group competition: moral norms are designed to promote cooperation and suppress competition.

While I take the evidence I've gathered to confer significantly plausibility on the hypotheses I've outlined, conclusively establishing that evolutionary debunking arguments have adequate empirical foundations requires a little extra work. We still need to consider to what extent the two hypotheses for which I've argued support the claims on which debunking arguments rely: that our moral psychology exhibits functional truth-irrelevance and/or phyletic contingency. These claims cannot be straightforwardly read off from the scientific work we've been considering in this chapter, as they involve a range of meta-ethical issues that cannot be addressed except by philosophical argumentation. That is the task of the next chapter.

2.

Interpreting the evidence:

Functional Truth-Irrelevance and Phyletic Contingency

1. Introduction

In the previous chapter, I presented evidence to support two evolutionary hypotheses about the origins of human morality: that morality is an outgrowth of social instincts inherited from the last common ancestor of *Homo* and *Pan*, and that moral norms have been adapted by group selection to suppress competition and promote cooperation. Building on these conclusions, this chapter considers to what extent the available evidence supports the key claims on which debunking arguments rely: that our moral psychology exhibits functional truth-irrelevance and/or phyletic contingency. I believe a convincing case can be made for both. Section 2 will clarify and then argue for the claim that our moral psychology exhibits functional truth-irrelevance; section 3 will do the same for the claim that our moral psychology exhibits phyletic contingency.

2. Functional truth-irrelevance

2.1 Introduction

The conditions of evolutionary success lead us to expect that animal minds will be adapted to reliably record fitness-relevant aspects of the environment: the senses must correctly identify and track objects and agents; memory must accurately record past events; updating on past evidence should reliably track real-world regularities. In the memorable words of Quine (1969): "Creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind." (126)

The general expectation, therefore, is that cognition is for truth-tracking.[1] According to the view that our moral psychology exhibits functional-truth irrelevance, moral cognition bucks this trend: although natural selection has shaped the character of human moral psychology, it has not done so for the sake of providing us with true beliefs about the right and good. This section argues in support of this claim. Section 2.2 seeks to clarify and situate the claim, and section 2.3 builds a sequence of interlocking considerations that I believe ought to convince us of its truth.

2.2 Clarifications and contrasts

2.2.1 Explaining the basic claim

The claim that our moral psychology exhibits functional truth-irrelevance can be stated as follows:

Functional Truth-Irrelevance:

Throughout human evolutionary history, selection pressures have favoured the evolution of (psychological structures dispositive of) certain moral beliefs, but there has not been selection for (psychological structures dispositive of) true moral beliefs.

It is important not to misinterpret this statement by confusing the *selection for/of* distinction.[2] The scope for confusion is well-illustrated by Sharon Street's (2006) treatment of the issue.

Street argues against what she calls the *Tracking Account*, but she varies in her definition of this position. At one point, she defines the *Tracking Account* as "the view that selective pressures pushed us toward the acceptance of the independent evaluative

[1] See Boulter (2007), Carruthers (1992), Fodor (1981), Millikan (1984a), Ruse (2006), Stephens (2001), Stewart-Williams (2005), Wilkins & Griffiths (2013). Plantinga (1993) famously denies this; for replies see Fales (2002), Fodor (2002), Ramsey (2002).
[2] Cf. Brosnan (2011).

truths" (135). Elsewhere, she offers a different gloss: "According to the tracking account … making such evaluative judgements [as we have evolved to make] contributed to reproductive success because they are *true*." (128) It is specifically the view on which the selective advantages conferred by certain evaluative dispositions are explained by the truth of corresponding ethical beliefs that Street argues against. Street evidently believes that the formulations I've just quoted describe the same position. In fact, her formulations elide the selection-for/of distinction.

A trait has been selected for iff it has evolved due to natural selection because it has raised the relative fitness of its bearers; a trait has been selection of iff it has evolved due to natural selection but this is *not* due to its beneficial effects on fitness. *Functional Truth-Irrelevance* says that there has not been selection for true moral beliefs (or psychological structures dispositive of the same): where selection has favoured some moral belief (or the disposition to accept it), this is not explained by the truth of the belief in question. *Functional Truth-Irrelevance* does not entail that (psychological structures dispositive of) true moral beliefs have not been selected of. To rule out that certain moral beliefs have been favoured by selection because they are true does not rule out that they are true: it merely denies the explanatory significance of this. For this reason, *Functional Truth-Irrelevance* does not straightforwardly rule out "that selective pressures pushed us toward the acceptance of the independent evaluative truths." It merely rules out that "making such evaluative judgements contributed to reproductive success because they are *true*." You can consistently affirm the former and deny the latter.

2.2.2 Meta-ethical connections: the moral explanations debate

Functional Truth-Irrelevance asserts that moral facts are irrelevant when it comes to explaining why certain moral beliefs have been favoured by selection. As noted by a

number of authors,[3] this suggests a connection to the *moral explanations debate* initiated by Gilbert Harman (1977, 1986) and Nicholas Sturgeon (1985, 1986). As I'll now explain, I think the connection is not as straightforward or illuminating as it might first appear.

Harman (1977) asks us to compare and contrast the following cases:

Vapour Trail:

A physicist observes a vapour trail in a cloud chamber. She makes the observational judgment that a proton is passing through the chamber.

Burning Cat:

You see some children pour gasoline on a cat and ignite it. You immediately form the observational judgment that what they are doing is wrong.

According to Harman, whereas it's reasonable to assume that a proton is present when explaining the physicist's judgment in *Vapour Trail*, an assumption about the wrongness of the children's action is explanatorily idle in *Burning Cat*: "an assumption about moral facts would seem to be totally irrelevant to the explanation of your making the judgment you make." (7) Harman subsequently tempers this verdict by suggesting that moral facts might be explanatorily relevant "if these facts were reducible in some way or other to other facts of a sort that might explain observations." (21)

Sturgeon (1985) rejects Harman's skepticism about the explanatory power of (unreduced) moral facts. He points to a variety of intuitively acceptable moral explanations: *e.g.*, that opposition to slavery rose in the period leading up to the American Civil War because slavery grew more oppressive during this time. Sturgeon

[3] *E.g*, Joyce (2006), Wielenberg (2010).

also argues that the wrongness of the action described in *Burning Cat* does explain your verdict, as you would not have judged the action wrong had it not been wrong: in that case, the children would not have been abusing the cat, and you would not have formed the verdict.[4]

This initial exchange has spawned a large body of work exploring the explanatory power (or lack thereof) of moral facts. Many philosophers have accepted Harman's position,[5] whereas others have joined Sturgeon in rejecting it.[6] It's beyond my remit to summarize the details of this on-going discussion here. Instead, I want to consider just how *Functional Truth-Irrelevance* relates to the lines already drawn in this debate. The key thing to note concerns the distinction between proximate and ultimate explanatory factors.

Although both Harman's discussion and *Functional Truth-Irrelevance* are in a broad sense concerned with whether moral facts figure in the explanation of human moral beliefs, the *Burning Cat* example chosen by Harman really asks us to consider whether moral facts are explanatorily relevant at the level of proximate causation. By contrast, *Functional Truth-Irrelevance* concerns the explanatory relevance of moral facts at the level of ultimate causes: specifically, in terms of functional explanation. *Functional Truth-Irrelevance* asks us to consider why individuals or societies upholding certain moral beliefs would have been favoured by natural selection over those upholding different moral beliefs or no beliefs at all. In an important sense, *Functional Truth-Irrelevance* concerns the explanation of events that are *downstream* from the adoption of moral beliefs. It does not address how our ancestors actually came to form such beliefs in the course of their lives, but rather why they were more successful than others once they had done so.

---

[4] This counterfactual criterion of explanatory significance has been widely criticised: see Harman (1986), Pust (2001a), Thomson (1996).
[5] See Audi (1997a), Blackburn (1984), Dworkin (1996), Leiter (2001), McGinn (1997), Quinn (1986), Thomson (1996), Williams (1986), Zangwill (2006), Zimmerman (1985).
[6] See Brink (1989), Cuneo (2006), Majors (2003), Oddie (2005), Wedgwood (2007), Wright (1992).

There is thus an important difference in focus between the questions about the explanatory power of moral facts raised in evolutionary debunking arguments and those central to the moral explanations debate. Someone could well accept the view that moral facts *are* relevant to explaining one's moral verdict in terms of immediate causation in *Burning Cat* whilst accepting *Functional Truth-Irrelevance*. The questions raised concerning the explanatory role of moral facts are quite different in each case.

2.3 The Case for Functional Truth-Irrelevance

Having thus clarified the nature of the claim, this subsection builds a case for *Functional Truth-Irrelevance* by drawing on the evolutionary hypotheses for which I argued in the previous chapter.

2.3.1 Functional Truth-Irrelevance and innate biases

In the previous chapter I argued firstly that our moral beliefs are influenced by innate biases inherited from the last common ancestor of *Homo* and *Pan*. It's easy to see that these biases did not evolve in order to incline us to believe moral truths. The moral judgments which they bias came about much later: since human beings are the only extant animal with moral beliefs, we should be highly confident that our last common ancestor did not hold any beliefs about right and wrong. Darwinian evolution does not 'look ahead': it follows no preformed, progressive trajectory. Thus, it did not anticipate that human beings would later form moral beliefs and pre-evolve a set of truth-conducive biases several million years in advance.

2.3.2 Functional Truth-Irrelevance and morality as a group-level adaptation

The fact that our moral outlook is an outgrowth of social instincts that did not evolve for the purpose of biasing true moral beliefs already provides some degree of support

for *Functional Truth-Irrelevance*. This view is further upheld when we consider the subsequent selection-pressures shaping human moral psychology.

The second claim for which I argued in the previous chapter is that the moral norms characteristic of human societies are adaptations favoured by group competition, designed to suppress competition and promote cooperation by altering the balance of costs and benefits associated with prosocial and antisocial behaviours. I offered two 'case studies' to illustrate this principle: the role of moral norms in coordinating punishment of free riders and driving up pay-offs in public goods settings, and the role of egalitarian norms in levelling fitness, supressing intragroup selection and promoting group-beneficial behaviours. This view of moral norms as adaptations supports *Functional Truth-Irrelevance*, I suggest, as it provides no apparent role for moral facts in explaining why certain moral beliefs have been favoured by group selection.

I didn't at any point in the previous chapter have to invoke any moral claim in explaining why certain moral beliefs have been selected for. The account I offered was value-free. For example, in explaining why selection favours norms among hunter-gatherer societies that require individuals to share meat, I did not have to say whether sharing meat is in fact morally required or morally praiseworthy. I had only to refer to the role of moral beliefs in motivating and coordinating criticism of those who do not share within the group, the threat of criticism in sustaining cooperation, and the benefits of pooling risk with respect to game kills. The ability of norms to reflect people's genuine obligations was nowhere at issue.

The claim that moral facts are not needed in adaptationist explanations does not suffice to establish *Functional Truth-Irrelevance*, however.[7] As noted in my

[7] Some authors may give this impression. For example, Street writes: "the tracking account is scientifically indefensible. To explain why human beings tend to make the normative judgments we do, we do not need to suppose that these judgments are *true*." (2008a: 209)

introduction, explanatory questions typically admit of multiple non-competing answers. Context determines what degree and what kind of explanatory information is appropriate; because the explanatory factors leading up to any event are typically vast in number, any successful explanation is selective to some extent. There is a sense in which nothing is ever needed to explain anything: given the right context, almost anything can be omitted. Our ability to explain the selection of certain moral beliefs without invoking corresponding moral facts is thus insufficient as an argument for *Functional Truth-Irrelevance* unless we can make a case that this reflects something over and above a license to disregard certain factors.

There is no better route to contextual relevance than explicit mention, and so a natural way to address the worry raised in the previous paragraph is to consider what explanatory mileage we can get by explicitly invoking moral facts in the context of explaining moral norms by selection. *Prima facie*, the answer is: none.[8] If we supplement the explanation already offered for the selection of meat-sharing norms with the claim that hunters ought in fact to share their kill throughout the group, we do not seem to contribute in any way to the explanation of why selection favours corresponding norms. The mentioned fact appears to have no relevance.

It might be replied that this merely reflects our ignorance. A philosopher attracted to some form of *Non-Analytic Reductive Naturalism* might well take this line.[9] According to this view, there exist true identity statements involving non-synonymous evaluative and non-evaluative predicates, analogous to theoretical identity statements in the natural sciences (*e.g.*, 'Water is H_2O'). In some cases, the explanatory significance of certain facts cannot be understood unless we are aware of relevant reductionist identities. For example, someone who does not understand that a gas's temperature is

[8] Cf. Street (2006: 129-132).
[9] For this meta-ethical view see, *inter alia*, Boyd (1988), Copp (1995, 2007, 2009), Jackson (1998), Schroeder (2007).

reducible to the kinetic energy of its molecules will not be able to understand how a rise in temperature within a gas explains the increased pressure on its container. Likewise, it might be suggested that our inability to see any potential relevance for moral facts in explaining moral norms by natural selection should dissolve once we accept some suitable reductionist account: for example, one that allows us to identify moral facts with facts about which norms best serve social aims related to cooperation and coordination.[10] This parallels Harman's claim, noted earlier, that the explanatory potency of moral facts can in principle be rescued by some form of naturalistic reduction. Richard Joyce (2006) notes this point in his discussion of evolutionary debunking arguments. To establish that moral facts are explanatorily irrelevant with respect to natural selection, Joyce considers it necessary to argue against *Reductive Naturalism.*

Like Joyce and many others, I count myself as skeptical of *Non-Analytic Reductive Naturalism.* This position is subject to a range of well-known problems. Thought-experiments involving *Moral Twin Earth* suggest that the semantics of moral terms are disanalogous to those of theoretical terms in the natural sciences in respect of the properties that generate the possibility of non-analytic reductive identity statements for the latter.[11] Among the metaphysical problems for *Non-Analytic Reductive Naturalism* we can appeal to the fact that the moral facts appear to *depend on* the non-moral facts:[12] the relation of dependence appears asymmetric and must therefore obtain among metaphysically distinct sets of facts.[13] We can also point out that moral facts are a species of normative fact, but that the normative is a broader genus, also comprising epistemic facts about what we ought to believe, demands of prudential rationality, and so on. The diversity of normative facts makes it difficult to

[10] For this sort of view, see, in particular, Boyd (1988), Carruthers & James (2008), Copp (1995, 2007, 2009).
[11] Horgan & Timmons (1991, 2000).
[12] Dancy (2004); Mackie (1977).
[13] For this argument, see Audi (1997a), McNaughton & Rawling (2003), Parfit (2011: 300-301).

see what all these facts could have in common except some irreducible element of shared normativity.[14] For example, epistemic facts do not appear, on the face of it, to have anything to do with social aims related to cooperation and coordination.

This, of course, is just a telegraphic list of objections. Each of these objections has received replies at one time or another. On this front David Copp (2007, 2009, 2012) has done a formidable job in defending *Non-Analytic Reductive Naturalism*. I don't mean to suggest that my listing these problems constitutes a decisive objection: there are epicycles upon epicycles still to go through. Sadly, they are beyond the scope of this essay. I merely want to register that *Non-Analytic Reductive Naturalism* is subject to a range of serious objections. The attempt to resist inferring *Functional Truth-Irrelevance* by appeal to *Non-Analytic Reductive Naturalism* is at least as problematic.

Ultimately, however, I don't believe it necessary to refute *Non-Analytic Reductive Naturalism* in order to secure a case for *Functional Truth-Irrelevance*. There is another route to this conclusion that allows us to bypass these concerns altogether.

2.3.3 Functional Truth-Irrelevance and our error-prone moral psychology

We can argue for *Functional Truth-Irrelevance* by drawing attention to the error-prone nature of our evolved moral psychology: our moral thinking is simply too flawed to sustain the hypothesis of selection for accuracy. A high frequency of errors associated with our evolved moral psychology is difficult to square with the assumption of an organ designed for accuracy whatever one's meta-ethics, so we can cut through any controversy about *Reductive Naturalism* via this route.

We can point to two kinds of error that support *Functional Truth-Irrelevance*. On the one hand, there are procedural errors in moral thinking. Our moral reasoning is often biased in ways that appear poorly designed relative to the aim of generating

[14] Cf. Parfit (2011: 363-364).

accurate moral beliefs. We can rely here on the *Social Intuitionist Model* of moral judgment developed by Haidt and the growing body of evidence in its support.[15] According to the model, our moral judgments are typically the result of immediate, affect-laden intuitions. The model doesn't deny that moral reasoning occurs, nor even that it can lead us to overrule our intuitions. However, it posits that the reasoning-process is typically not free to search for the truth, but enlisted to construct considerations that support pre-formed judgments. Reason is, as it were, the slave of the passions.

This tendency is most clear in our propensity for confabulation: we readily invent spurious reasons when pressed to justify pre-formed moral conclusions. Subjects asked to explain why a hypothetical case of sibling incest is wrong will reliably cite the risk of producing a child with birth defects even if the vignette describing the case explicitly rules out this possibility, detailing the use of multiple forms of birth control.[16] When people are tricked into believing they assented to a moral position with which they actually disagreed moments ago, they readily begin to construct coherent justifications in its support.[17] People's political positions are dictated by their feelings of liking for politicians or parties rather than their factual knowledge of relevant issues,[18] and they are often resistant to updating their beliefs in light of corrective factual information.[19] Defining *epistemic functionalism* as the view that "moral thinking is performed *in order to* find moral truth," (808) Haidt and Selin Kesebir (2010) conclude: "The many biases, hypocrisies, and outrageous conclusions of (other) people's moral thinking are hard to explain from an epistemic functionalist perspective" (814).

What about the intuitions themselves? According to Haidt, the legacy of our

[15] See Haidt (2001, 2007), Haidt & Bjorklund (2008), Haidt & Kesebir (2010).
[16] Haidt et al. (2000).
[17] Hall et al. (2012).
[18] Westen (2007).
[19] Kuklinski (2000); Lord et al. (1979); Nyhan & Reifler (2010).

evolution is felt primarily in the emotive intuitional judgments that people deliver automatically and without conscious reasoning.[20] Besides the procedural errors just noted, we can point to a prevalence of substantive errors in our evolved moral psychology: false conclusions, rather than mere bad reasoning. This point is emphasized by Street (2006) with respect to in-group bias. As she notes, human beings have a deep tendency to believe that members of other ethnic groups deserve lesser treatment. This is readily explained by the adaptationist hypothesis for which I argued in the previous chapter, with its emphasis on group competition. Most philosophers would nonetheless staunchly deny that ethnic boundaries are morally relevant in this way. As Street (2006: 133) notes, this entrenched error is difficult to explain if we assume selection for true moral beliefs. We can carry this line of argument further, relying on somewhat more controversial moral opinions, albeit ones that most philosophers would be expected to endorse. Recall that according to Haidt's *Moral Foundations Theory*, the moral mind is equipped with five innate modules, each of which predisposes us to moralize a distinct domain of social action via corresponding intuitions. Each module is supposed to represent a psychological adaptation to a particular aspect of social life that held high importance for the fitness of our ancestors. However, not everyone makes equal use of these foundations. Political liberals are said to do almost the entirety of their moral thinking by reference to the first two domains, regarding the others as sources of prejudice and moral backwardness.[21] Evidence suggests that the majority of academic philosophers should fall into this category, as professors in the humanities and social sciences are overwhelmingly liberal.[22] Liberals should view three out of five evolved modules as being primarily sources of distorted moral conclusions. This degree of unreliability is difficult to explain on the view on which there has been selection for true moral beliefs.

[20] See Haidt (2001), Haidt & Bjorklund (2008).
[21] Graham et al. (2009); Haidt (2012).
[22] Gross & Simmons (2007).

2.3.4 Summary

A range of considerations thus lend support to *Functional Truth-Irrelevance*. Our moral outlook is an outgrowth of social instincts that did not evolve to facilitate accurate moral beliefs. The subsequent selection-pressures shaping our moral psychology can be explained without appeal to the correctness of relevant norms, and mention of correctness appears to offer no added explanatory insight. Whereas this last point might be said by proponents of *Non-Analytic Reductive Naturalism* to reflect mere ignorance, there are multiple reasons to be skeptical of this meta-ethical position. Finally, the view that there has been selection for true moral beliefs is difficult to square with the observation of wide-spread errors, both procedural and substantive. All in all, we have considerable reason to suppose that the moral mind was not designed for the production of true beliefs.

3. Phyletic contingency

3.1 Introduction

The traditional Creationist view designates human beings as the pinnacle and purpose of the biological world. Even as naturalists in the 19th century came to realize that humanity is the product of evolution, the idea quickly took hold that the evolutionary process followed a directed trajectory, aiming all along at the production of *Homo sapiens*.[23] By contrast, the Darwinian understanding of evolution implies that our species enjoys no such privileged position: our emergence in the African Rift Valley some 200,000 years ago was neither inevitable - nor even probable - when viewed through the lens of natural history. In the memorable words of Stephen Jay Gould (1990), we are "a tiny twig on an improbable branch of a contingent limb on a

[23] See Bowler (1989).

fortunate tree" (291).

E. O. Wilson (2004) reminds us that we also should not mistake the quirks of our species for inevitable concomitants of advanced intelligence and complex social structure. As Wilson puts it: "Civilisation is not intrinsically linked to hominoids. Only by accident was it linked to the anatomy of bare-skinned, bipedal mammals and the peculiar qualities of human nature." (23) The suggestion that our moral outlook exhibits phyletic contingency asks us, in a similar vein, to give up pretensions of biological inevitability in the moral rules characteristically endorsed by human societies. We are asked to see these rules as reflecting contingencies of our evolution and species-nature. That human morality is the only morality is no less arbitrary or contingent than our current monopoly on advanced intelligence.

A number of prominent evolutionary scientists have taken up this view, beginning with Darwin and his remarks on the moral norms that should be expected of intelligent bees. Writing with Michael Ruse, Wilson himself would go on to suggest that "[i]t is easy to conceive of an alien intelligent species evolving rules its members consider highly moral but which are repugnant to human beings" and that "ethical premises are the peculiar products of genetic history" (Ruse & Wilson 1986: 186). Most recently, Herbert Gintis (2013) has put forward a view of this kind about intuitions and beliefs concerned with property rights. Gintis takes these beliefs to rest on innate biases, as many animal species (including closely related nonhuman primates) exhibit an innate disposition to recognize and respect prior possession for items including food and territory. Although this disposition is widely distributed across taxa, Gintis extracts a moral of biological contingency. He writes:

> modern notions of property are built on human behavioral propensities that we share with many nonhuman animals. Doubtless, an alien species with a genetic organization akin to our ants or termites would find our notions of individuality and privacy curious at best, and probably incomprehensible.

(127)

One peculiar feature of these claims deserves mention. Evolutionary biology is, at its heart, a historical enterprise devoted to explaining the current lay of the biological land by appeal to its past. The claims I've just quoted, however, are in effect predictions about how the evolutionary process will unfold under certain conditions. Natural history is often felt to be unpredictable in the extreme: too much is down to chance and randomness.[24] Thus, the attempt to predict what might have emerged from some branch on the tree of life given favourable counterfactual conditions may seem more like science-fiction than science. Gould (1990) writes: "we cannot even make predictions when we know the line of descent: we cannot see the mayfly in *Aysheiaia*, or the black widow spider in *Sanctacaris*. How can we specify the world that different decimations would have produced?" (292)

We do well to take on board the point that attempts to predict the course of natural history are speculative and uncertain in comparison to our efforts to explain its past trajectory. However, we also shouldn't exaggerate our epistemic disadvantage. Gould (1990: 289) himself suggests that many important facets of natural history exhibit something resembling inevitability: for example, he suggests that mobile multicellular organisms built by cell division are expected to exhibit bilateral symmetry in body plan. I think that we can reasonably make similar predictions about the fit of moral norms to the social life of the species they uphold. In this section I present a series of considerations that I believe confer significant plausibility on predictions like those advanced by Darwin, Wilson, and Gintis. Our moral psychology, I argue, bears the hallmarks of our peculiar place within the tree of life.

3.2 Clarifications and contrasts

[24] See *esp.* Beatty (1995).

3.2.1 Explaining the basic claim

The basic claim that our moral psychology exhibits phyletic contingency can be stated as follows:

Phyletic Contingency:

Had morality evolved in a distantly related species with a very different form of social organization, their moral beliefs would have been correspondingly different to those characteristic of human societies.

This statement is in need of clarification concerning the nature of the differences between our moral beliefs and those we could expect of distantly related species. In particular, we need to consider whether we should expect that other evolved moral systems would be not merely different, but different in such a way as to disagree with our own, and, if so, what kind of disagreement we might expect.

On one reading, we could accept *Phyletic Contingency* without any commitment to disagreement. A distantly related species with a different form of social organisation would no doubt have norms and beliefs about forms of social behaviour without equivalent in human societies. In that respect, their norms might be different from ours: they would concern issues not touched on in our moral thinking. *Phyletic Contingency* might be counted as true on that basis. I take it that no one is disturbed by the claim so interpreted. We are disturbed only if we are led to expect that other evolved moral systems would disagree with our own if applied to the same question(s).

The expectation of disagreement is necessary to intimate worries about debunking, though also insufficient of itself. Everything depends on the sort of disagreement that we expect. Philosophers are well-acquainted with the suggestion that the moral disagreements we observe in ordinary life don't reflect deep differences

in values, but rather underlying differences in beliefs about relevant non-moral questions and/or failures of epistemic rationality.[25] It's possible that other species might evolve moral beliefs that would disagree with our own because of such factors. Again, I take it nobody is disturbed by this possibility. We are disturbed if and to the extent that we are led to believe that our values depend fundamentally on contingencies of our evolutionary history: other evolved moral systems would disagree with our own if applied to the same question and this disagreement would not be traceable to differences in relevant beliefs about non-moral questions and/or failures of epistemic rationality.

One plausible means by which to develop this proposal is via the suggestion that our moral intuitions, in particular, are parochial to our phylogeny. Other species with other forms of social organisation would then be expected to have different moral intuitions, and these would be of a kind that would lead to disagreements not attributable to differences in relevant non-moral beliefs or failures of epistemic rationality, but simply to different ways of seeing things (so to speak). That is what we should expect if their moral psychology is otherwise like ours except for those contrasts attributable to differences in phylogeny: as noted, our own evolved moral psychology appears to be driven primarily by immediate, affect-laden intuitions.

3.2.2 Meta-ethical connections: radical disagreement

Let's say that a moral disagreement not attributable to differences in relevant non-moral beliefs or failures of epistemic rationality is a *brute disagreement*. By interpreting *Phyletic Contingency* so as to refer to the possibility of brute disagreement, we may wonder how my discussion relates to more familiar meta-ethical debates about the significance of moral disagreement. As with the relation of *Functional Truth-*

[25] See, *inter alia*, Boyd (1988), Brink (1989: 197-209), Enoch (2009), Huemer (2005: 137-139), Shafer-Landau (2003: 218-220), Smith (1994: 187-189).

Irrelevance to the moral explanations debate, I think the connection here is not quite as straightforward and illuminating as it might first appear.

It is widely believed that moral disagreement is a problem for *Meta-Ethical Realism*, though there is little agreement on what sort of phenomenon involving disagreement is problematic, nor which aspect of *Meta-Ethical Realism* is thereby challenged. Sometimes the problem concerns actual disagreement – in particular, its prevalence and apparent entrenchment.[26] In other cases, the problem for *Realism* is supposed to arise from the mere possibility of certain forms of radical moral disagreement.[27] The anti-realist implication is sometimes thought to be semantic in character: phenomena associated with disagreement have been taken to show that moral sentences do not purport to describe moral facts.[28] In other cases, the anti-realist implication is more straightforwardly metaphysical: that there are no moral facts[29] or no moral facts of the metaphysically robust sort posited by *Meta-Ethical Realism.*[30]

By appealing to the possibility of brute moral disagreement, we may worry that I can't avoid prejudging certain aspects of this debate: I have to take sides against either *Realism* or those who believe the possibility of radical moral disagreement is incompatible with *Realism.* In reality, I can afford to be more ecumenical than this. The definition of brute disagreement I've adopted differs importantly from the kind of radical disagreement whose possibility is the target of meta-ethical debate in this area. As Folke Tersman (2006) notes: "The alleged difference between ethics and areas that should be construed realistically is that many disagreements that arise within ethics cannot be rationally resolved." (30) A brute disagreement, on my understanding, is a

[26] Harman (1996); Mackie (1977).
[27] Schiffer (2003); Wright (1992, 1995).
[28] Hare (1952); Harman (1996); Tersman (2006); Wong (1984).
[29] Mackie (1977); Schiffer (2003).
[30] Wright (1992, 1995).

disagreement that is not traceable to underlying differences in beliefs about relevant non-moral questions and/or failures of epistemic rationality: for example, a disagreement that rests on a brute difference in intuitions. This does *not* imply that a brute disagreement is irresolvable in principle or that it will persist throughout any possible epistemic improvements. It does not rule out that one side may later be able to discover additional considerations that decide the issue in her favour or a genealogical debunking argument that selectively undermines the opposing view.

Many writers in the meta-ethical debate appear to assume that a brute disagreement must be rationally irresolvable: questions about whether actual moral controversies are cases of radical moral disagreement are frequently taken to turn on whether people's varying beliefs can be explained by differences in relevant non-moral beliefs and/or failures of epistemic rationality.[31] My own experience is that careful reflection is frequently able to turn up novel considerations capable of deciding difficult issues on which people otherwise have nothing to go on but powerful, divergent intuitions.[32] As Parfit (1984: 453) reminds us, we should be careful not to prejudge the progress that may be achievable through the careful practice of moral philosophy, because non-religious ethics is young and under-developed. Whether any particular brute disagreement is rationally resolvable or irresolvable is, I think, an open question.

It does not seem especially plausible that *Meta-Ethical Realism* is challenged by the possibility of a brute disagreement that is *also* rationally resolvable. Since *Phyletic Contingency* makes no appeal to the possibility of disagreements that are irresolvable in principle, it does not touch on the possibility of radical disagreements of the kind that really matter in traditional meta-ethical debates, much as participants in these debates have often assumed that brute disagreements must be radical disagreements.

---

[31] *E.g.*, by Brandt (1954, 1959), Doris & Plakias (2008), Doris & Stich (2005), Wright (1992).
[32] Key examples include Broome (2005), Singer (1976), Thomson (1971).

3.3 Arguing for Phyletic Contingency

This subsection builds a case for *Phyletic Contingency*, drawing on the evolutionary hypotheses for which I argued in the last chapter, as well as evidence that brute disagreement already occurs across human societies.

3.3.1 Phyletic Contingency and innate biases

I argued previously that human morality is an outgrowth of affective dispositions inherited from the last common ancestor of *Homo* and *Pan*. The fact that human morality is continuous with a form of 'proto-morality' attributable to a distant primate ancestor of itself provides a fair degree of support for *Phyletic Contingency*. Our moral concerns reflect a way of life and a set of congruent emotional responses shared with similar African apes. We should expect, therefore, that moral systems evolving in distantly related species would be tailored to reflect corresponding differences in phylogeny.

3.3.2 Phyletic Contingency and morality as a group-level adaptation

We can generate further support for this view by considering the picture of morality as a group-level adaptation for which I argued in the previous chapter. The selection pressures shaping human morality provide no grounds on which to expect convergence and quite a few that suggest the possibility of disagreement.

On the one hand, *Functional Truth-Irrelevance* rules out any prospect of convergence based on selection for true moral beliefs. Even to the extent that we suppose our own evolved moral beliefs to be correct, this provides no expectation that selection will not favour conflicting moral systems under counterfactual conditions. Falsehood is no impediment to usefulness.

In addition, according to the proposal for which I argued in the previous chapter, morality exists to promote cooperation and suppress competition. Opportunities for cooperation and conflict are likely to differ greatly across taxa. For example, I argued that moral rules found across human societies can be seen as attempts to manage an innate disposition for dominance-seeking behaviour. Dominance-seeking is by no means a biological universal in social species: social species can be plotted along a continuum between 'egalitarian' and 'despotic',[33] and in some primates, such as squirrel monkeys, we see little intragroup dominance.

Of itself, the fact that moral rules should be expected to be adapted to different forms of competition and cooperation gives little support to anything over and above the suggestion that norms might differ across taxa in terms of their subject-matter or focus. There is no reason, however, to expect that our moral intuitions and beliefs are constrained in their application by the contingent conditions of their evolutionary usefulness. For example, moral condemnation of incest is not limited to cases that involve a risk of producing offspring with genetic diseases: as noted earlier, sibling incest is condemned even if multiple forms of birth control are used. Similarly, people's intuitions about punishment are resolutely retributivist.[34] Retributivism, by definition, is a view which prescribes punishment even when punishment has no utility. Because our evolved moral beliefs don't come tagged with the conditions of their own evolutionary usefulness, we should expect that they are liable to conflict with those of species whose norms are adapted to a different set of social conditions, requiring a correspondingly divergent set of norms. This disagreement need not be traceable to differences in beliefs about those aspects of the environment which have differed across the two species, any more than human beings' condemnation of incest is driven in the first instance by the belief that there is a risk of producing offspring with innate

[33] Vehrencamp (1983).
[34] See Carlsmith et al. (2002).

defects.

3.3.3 Brute disagreement across cultures

A final consideration that supports *Phyletic Contingency* is that we already appear to observe, within our own species, brute disagreements reflecting underlying differences in culture and social organisation. Much as we're impressed by the extent of human cultural diversity, from a biological perspective human societies are quite limited in their variability. As Wilson (2004) notes: "all versions of human social behavior together form only a tiny fraction of the realized organizations of social species on this planet and a still smaller fraction of those that can be readily imagined with the aid of sociobiological theory." (18-19) We should expect, therefore, that the possibilities for brute disagreement are even greater when considering the moral beliefs that might emerge in distantly related species with markedly different forms of social organisation.

It is a commonplace that anthropological research shows moral norms to differ greatly across human cultures. This view was held long before Geertz and Evans-Pritchard made the discovery of difference the central mission of ethnography. Locke (1689/1975) writes that "there is scarce that Principle of Morality to be named, or *Rule of Vertue* to be thought on … which is not, somewhere or other, *slighted* and condemned by the general Fashion of *whole Societies* of Men" (72). A commonly held view is that ethnographic research nonetheless typically fails to provide sufficient detail about the moral beliefs of distant cultures to determine whether divergent moral belief reflect brute disagreements: for example, the information provided may be too shallow to eliminate the possibility that divergent moral norms are associated with differences in relevant non-moral beliefs.[35] Even so, we are able to pick out some

[35] See Abarbanell & Hauser (2010), Brandt (1959), Doris & Plakias (2008), Moody-Adams (1997).

relatively clear cases on which a verdict of brute disagreement can be returned with reasonable confidence.[36] I'll discuss two such cases in the following, and then offer a general argument to suggest that these are unlikely to be exceptional.

I begin with Richard Brandt's (1954, 1959) well-known work on Hopi ethics. It was an explicit aim of Brandt's ethnography to determine whether differences in moral norms between the Hopi and majority Anglophone culture reflect brute disagreements. Brandt argued that a brute disagreement occurs in respect of attitudes toward animal cruelty. He found that Hopi culture was generally more permissive toward the infliction of animal suffering. Hopi children would capture birds, tie them with string, and 'play' with them: "This play is rough, and birds seldom survive for long." (1954: 213) Young men would also engage in a game known as 'chicken pull'. This involves burying a live chicken up to its neck, after which players would attempt to pull the bird out of the ground from horseback. "When someone succeeds in this, the idea is then for the other contestants to take away from him as much of the chicken as they can. The 'winner' is the one who ends up with the most chicken." (1959: 102) Brandt sought to identify underlying differences in beliefs about the nature of animals that might account for the greater permissiveness of Hopi culture in regard to animal cruelty. Instead, he found that the Hopi frequently held beliefs about animals of a kind that would lead us to expect more restrictive attitudes on their part. For example, one of Brandt's informants was a young man who enjoyed playing 'chicken pull'; when this man's horse broke a leg, he had made no effort to euthanize it, instead allowing the animal to run off and die on the mesa. The same young man told Brandt: "Dogs or horses have sense as much as a human being, yet they can't talk. If they could speak, they'd be as smart as humans. … Animals have pain as much as human flesh, but they won't cry out or say anything." (1954: 215) In general, Brandt argued, the Hopi were

[36] See Doris & Plakias (2008), Doris & Stich (2005) to which my discussion here is indebted.

more likely than members of the majority Anglophone culture to see animals as similar to human beings. Nonetheless, they were more likely to accord animals lower moral status. Brandt (1954) concluded that, with respect to animal cruelty, "there is a basic difference of attitude between Hopi and white Americans" (245).

As a second example, we'll consider a recent set of experiments performed by Linda Abarbanell and Marc Hauser (2010) involving the moral intuitions of rural Mayans, identified by the authors as a plausible candidate for brute disagreement. These experiments required subjects to deliver intuitive verdicts in response to a series of moral dilemmas broadly similar to the *Trolley Case* and *Footbridge Case*. The dilemmas had previously been trialled in large, web-based surveys, showing that subjects' intuitions conformed to ethical principles familiar from deontological moral theory: participants attributed greater wrongness to harms arising from action as opposed to omission and to harms arising from intention as opposed to foresight.

Abarbanell and Hauser put similar materials to a group of Mayan subjects living in small-scale societies based on slash-and-burn agriculture in Chiapas, Mexico. The judgments delivered by rural Mayans were like those previously observed in respect of attributing greater wrongness to harms arising from intention as opposed to foresight. However, they did not attribute greater wrongness to harms resulting from action as opposed to omission. Abarbanell and Hauser sought to test whether Mayan subjects might regard acts and omissions as equivalent in terms of causal responsibility for harm, but found that participants did attribute greater causal responsibility when harms were caused by action as opposed to omission: they simply attached no intuitive moral significance to this. Notably, a further experiment involving urbanized Mayans found that they were unlike their rural counterparts and more like participants in web-based surveys: they attributed greater wrongness to harms arising from action as opposed to omission. Thus, Abarbanell and Hauser infer

that indifference to the acts/omissions distinction probably reflects some aspect of the close-knit character of rural, agrarian society.

Some might hold out for the possibility that there is something more than a brute difference in intuition at work in these examples.[37] Obviously, it's difficult to prove a negative. However, there are certain general considerations which suggest that brute disagreement is to be expected. These considerations appeal to the nature of the internal proximate causes of moral judgment, relying once more on the *Social Intuitionist Model*. According to the model, moral judgments are frequently arrived at on the basis of immediate, affect-laden intuitions; reasoning from general principles is often *post hoc* and aimed at confirming one's pre-established judgment. Evidence of the kind reviewed in the previous chapter suggests, furthermore, that the intuitive responses of children and adults are malleable by way of framing effects and techniques designed to induce emotion.[38] This suggests that disagreements resting on differences of intuition are to be expected.[39] Where moral disagreements occur, they may well be the rule, rather than the exception. Furthermore, evidence which might otherwise be thought to disconfirm the existence of brute disagreement should be handled with some degree of skepticism. Where we find that those who disagree morally also hold divergent non-moral beliefs of a kind that might rationalize conflicting moral verdicts, we should not assume without further argument that causation runs from the latter to the former and that disagreement would evaporate in the face of convergent non-moral beliefs.

Given the plausibility of the view that differences among human societies generate brute disagreements about moral issues, we should expect that the scope for disagreement is only intensified when considering the possible moral systems that

[37] For discussion see Moody-Adams (1997) and Fraser & Hauser (2010).
[38] Cf. Sinnott-Armstrong (2006b, 2011).
[39] Cf. Mackie (1977: 37-38).

might evolve among distantly related species. As I noted in the previous chapter, the variability of human moral codes has to be seen against a background of 'thematic clustering', with high-level themes corresponding to Haidt's *Moral Foundations* observable across all (or almost all) societies. Taken in conjunction with the evidence that our moral concerns reflect emotional responses shared with similar African apes and that our moral norms have evolved to respond to adaptive challenges particular to our subfamily within the primate order, it is quite implausible that our species has exhausted the bounds of moral diversity and brute disagreement.

3.3.4 Summary

A range of considerations thus lend support to *Phyletic Contingency*. Our moral outlook is continuous with affective dispositions in closely related primate species, suggesting that human morality reflects the particular conditions of hominine social life. The subsequent selection pressures shaping our moral psychology are similarly contingent on idiosyncratic features of our phylogeny, such as our innate dispositions for dominance-seeking behaviour. Finally, we have evidence of brute disagreements within our own species contingent on differences of social organization and identity. All in all, we have considerable reason to suppose that human moral beliefs reflect the contingent conditions of our evolutionary descent.

4. Conclusion

This marks the end of the first major division in this essay, designed to supply the foundations for evolutionary debunking arguments. The previous chapter identified two well-supported evolutionary hypotheses about the origins of morality, and this chapter has built on these hypotheses to provide a case for *Functional Truth-Irrelevance* and *Phyletic Contingency*. The question for us now is whether we can build on these

foundations to argue that evolutionary explanations serve to debunk at least some of our moral beliefs. We are thus going to shift our attention to the core epistemological questions at the centre of the debate. My thesis is that the most popular arguments, which appeal to *Functional Truth-Irrelevance* but not *Phyletic Contingency*, are flawed and unpersuasive. The next two chapters are devoted to substantiating this position. A successful debunking argument needs to be anchored in *Phyletic Contingency*, I claim, and the remaining chapters of the essay will substantiate that a successful debunking argument can be built along these lines.

3.

Ockham's Razor, Sensitivity, and the Total/Functional Fallacy

1. Introduction

In this chapter and the next, we are going to consider a number of arguments that seek to derive debunking implications from *Functional Truth-Irrelevance* without appeal to *Phyletic Contingency*. These arguments are narrowly focused on the issue of the irrelevance of moral facts within functional explanations. They are the most prominent arguments in the literature today: the sort of arguments offered by Richard Joyce, Michael Ruse, and Sharon Street. I am going to try to convince you that these arguments fail.

In this chapter, I consider two lines of reasoning: the first relies on an appeal to *Ockham's Razor*, the second on considerations of Nozickian sensitivity. I note that both are subject to a number of epistemological problems. They also share a deeper flaw. They rely on what I call the *Total/Functional Fallacy*. This involves inferring that moral facts are totally irrelevant in explaining certain elements of our moral psychology from the premise that they are irrelevant in corresponding functional explanations. In other words, these arguments presume that since we can explain why certain moral beliefs/intuitions would have been selectively advantageous regardless of their truth or falsity, we should infer that moral facts are explanatorily irrelevant altogether in accounting for the relevant beliefs/intuitions. As I explain in section 4, this is a mistake.

Here is the plan for this chapter. In section 2, I explore how philosophers have appealed to *Ockham's Razor* in arguing that evolutionary explanations are debunking. I suggest that these appeals gain plausibility from an implicit epistemological assumption, *Classical Empiricism*, which is ultimately disastrous because it leads to external-world skepticism. In section 3, I consider debunking arguments that appeal to considerations of Nozickian sensitivity. The epistemological presuppositions required here also threaten to lead us into skepticism, but I suggest that there are not-

implausible means of blocking this implication. Epistemological problems to do with inductive inference are not so easily evaded, but their seriousness is not so clear. Having explored these epistemological issues, in section 4 I argue that even if these difficulties could be taken in stride, both lines of argument ought to be rejected because they rely on the *Total/Functional Fallacy*.

2. Ockham's Razor

2.1 Some examples

A number of debunking arguments appeal to *Ockham's Razor*, whether explicitly or implicitly. According to Richard Joyce (2006), once we understand how our moral judgments have evolved, "Ockham's Razor will leave us with no reason to believe in moral facts." (195) Moral facts, he claims, "should be excised from the picture with a swift slash of Ockham's Razor, since we have a complete explanation of moral judgment with no need to posit any extra ontology in the form of moral facts." (188) A similar line of reasoning is proposed by Alan Gibbard (1990) in arguing that evolutionary explanations undermine belief in normative facts. Having suggested that normative judgments have the evolved function of coordinating actions for mutual benefit, Gibbard writes:

> If the hypothesis holds good, we do not need normative facts to explain our making the normative judgments we do. Our making them is to be explained by the rewards of coordination. To suppose that there are normative facts is gratuitous. (107-108)

Similarly, Michael Ruse (1986) suggests that evolutionary explanations threaten belief in objective moral facts by showing that "the objective foundation for morality is redundant" (254). In a later paper, co-authored with Edward O. Wilson, he writes: "The evolutionary explanation makes the objective morality redundant ... And surely,

redundancy is the last predicate that an objective morality can possess." (Ruse & Wilson 1986: 186-187)[1]

2.2 Clarifying Ockham's Razor

In evaluating this line of argument, let's begin by clarifying what exactly *Ockham's Razor* requires of us. As E. C. Barnes (2000) notes, there are actually two distinct principles of parsimony referred to under this name.

The first of these Barnes calls the *Anti-Quantity Principle*. This says that theories with fewer theoretical components are *pro tanto* preferable. By a 'theoretical component' we mean any theoretical hypothesis positing entities, events, or processes, or ascribing properties to such. Here's a nice illustration of the *Anti-Quantity Principle*, drawn from the history of particle physics, which I borrow from Daniel Nolan (1997). In the early twentieth century, physicists were puzzled by the phenomenon of β decay. It was known that β decay involves the emission of an electron from the nucleus of a radioactive atom, but the drop in mass-energy following decay was greater than could be accounted for on this basis. An hypothesis originally due to Pauli, subsequently refined and popularized by Fermi, held that β decay involved the emission of a hitherto-unknown, neutrally-charged particle: the *neutrino*. This was to account for the missing mass-energy. As Nolan points out, the missing mass-energy could equally well have been accounted for by positing two emitted particles, each having a mass-energy of 1/2 the required amount, or seventeen million particles each carrying 1/17M of the missing mass-energy. The only thing favouring the postulate of a single neutrino is quantitative parsimony. A single particle is preferable according to the *Anti-Quantity Principle*.

The *Anti-Quantity Principle* is nonetheless subject to a good deal of

[1] See also Kitcher (2006), Sommers & Rosenberg (2003).

philosophical controversy. Richard Swinburne (1997) maintains that it's a brute and necessary fact that more parsimonious theories are more likely to be correct; Elliott Sober (1990) thinks a preference for more parsimonious theories must be legitimated by contingent, empirical considerations specific to a domain of inquiry. You'll be glad to know we don't have to take sides in this argument. The *Anti-Quantity Principle* is not the right formulation of *Ockham's Razor* for our purposes.

We are interested in what Barnes calls the *Anti-Superfluity Principle*. This requires us not to accept postulates which are explanatorily superfluous. It is the *Anti-Superfluity Principle* that is often thought to undermine belief in God given the existence of naturalistic accounts for the origins of life, the universe, and everything. It's clearly this principle that Joyce *et al.* have in mind. Stated as such, however, the *Anti-Superfluity Principle* is underspecified. Asked not to accept explanatorily superfluous postulates, we should ask: 'Explanatorily superfluous with respect to *what*?'

Suppose Joyce is right that moral facts are explanatorily superfluous when it comes to accounting for our moral beliefs/intuitions. This fails to establish that moral facts are explanatorily superfluous *tout court*. As David Brink (1989) notes, "if there were moral facts, they could be explanatory without explaining nonmoral facts; they could and would explain other *moral facts*." (183)[2] Moral facts might also explain nonmoral evaluative facts, even if they fail to explain any purely descriptive facts: Noël Carroll (1996) and Berys Gaut (2007) argue that the moral character of an artwork can determine it aesthetic properties. It's also worth noting that we seem perfectly justified in accepting certain postulates that are explanatorily superfluous with respect to our total evidence. For example, we are justified in postulating the existence of

[2] Cf. Shafer-Landau (2003: 104).

stars outside our light-cone.[3]

Barnes suggests that the *Anti-Superfluity Principle* should be restricted to cases in which the components of a theory are justified, if at all, by their explanatory power.[4] In cases in which we propose a theory *T* in order to explain a certain class of phenomena and the components of *T* should be assigned very low confidence except insofar as they contribute to the best available explanation of the phenomena in question, we are not justified in accepting any component of *T* unless that component contributes to the explanatory power of *T*. On this construal, the *Anti-Superfluity Principle* is more or less tautologous. That needn't be a point against it: tautologies have the lovely property of being necessarily true. This understanding of the principle does, however, raise some doubts about its applicability in the context of evolutionary debunking arguments. It's far from obvious that moral values are posits to be justified by their capacity to explain aspects of our moral psychology. Only under that assumption can evolutionary theories render our moral beliefs vulnerable to debunking arguments appealing to the *Anti-Superfluity Principle*. The next subsection explores this issue in greater depth.

2.3 Moral values as explanatory posits

No one would deny that moral claims are sometimes justified by appeal to their explanatory power. For example, someone might justify her acceptance of the *Doctrine of Double Effect* on the basis of its ability to explain why killing is permissible in the *Trolley Case* but not the *Footbridge Case*. But here the explananda are themselves moral propositions, not psychological phenomena. This is a matter of moral facts explaining

[3] Cf. Colyvan (1998).

[4] Of both the *Anti-Superfluity Principle* and the *Anti-Quantity Principle*, Barnes (2000) says: "these principles are intended only to apply to the evaluation of theories insofar as theories are supported by a body of data they purport to explain. Both principles are clearly not intended to apply to other cases of evidential support." (358)

moral facts.

Many philosophers explicitly reject the view that moral values are to be justified as explanatory posits that account for our moral judgments and intuitions. David Copp (1990) and Russ Shafer-Landau (2003: 110-114) both stress that morality is fundamentally a normative enterprise, and not an attempt to predict our observations in the manner of a scientific theory. Elliott Sober (1994) offers exactly this objection to the suggestion that evolutionary considerations might debunk our moral beliefs by rendering moral facts explanatorily redundant with respect to the phenomena of moral psychology. Considering the principle, "It is reasonable to postulate the existence of ethical facts only if that postulate is needed to explain why people have the ethical beliefs they do", Sober contends that it is "radically implausible" (108). He says: "Ethics is not psychology. The point of normative ethical statements is not to *describe* why we believe and act as we do, but to *guide* our thought and behavior." (109)[5]

Intuitively, I think most of us are inclined to agree with Sober's conception of the difference between ethics and psychology: the view that moral theories should be evaluated as psychological theories is something we'd have to be argued into. Nonetheless, some philosophers do propose to treat moral claims as explanatory posits of just this kind. For example, in constructing his naturalistic moral realism, Peter Railton (1986) endorses what he calls "the generic stratagem of naturalistic realism" (171), which he explains as follows: "to postulate a realm of facts in virtue of the contribution they would make to the *a posteriori* explanation of certain features of our experience." (171-172) By means of his reforming naturalistic definitions, Railton offers an account of the nature of moral value on which the stratagem is supposed to license belief in moral facts.

[5] Sober (2012) re-iterates this point.

Railton originally illustrates the stratagem by reference to our acceptance of claims about ordinary, medium-sized objects: chairs, tables, hands, arms, *etc.* He writes: "an external world is posited to explain the coherence, stability, and intersubjectivity of sense-experience. A moral realist who would avail himself of this stratagem must show that the postulation of moral facts similarly can have an explanatory function." (172) Thus, when Railton speaks of the contribution that some realm of fact must make with respect to explaining "certain features of experience", he does not have in mind those features of the external world which register in perception, but rather the private mental episodes which constitute their registering. Railton is here endorsing the programme for ontology set out by Quine (1948, 1951). Quine (1948) writes:

> Our acceptance of an ontology is, I think, similar in principle to our acceptance of a scientific theory, say a system of physics; we adopt, at least insofar as we are reasonable, the simplest conceptual scheme into which the disordered fragments of raw experience can be fitted and arranged. (35-36)

On this conception, the entire edifice of human knowledge is, at heart, a means for predicting and explaining features of our psychology, principally the nature of sense-experience. Quine and Railton are relying on the traditional empiricist assumption that our beliefs about the external world must be capable of being supported by inferences from premises describing the character of subjective experience. As Quine puts it: "Bodies are not given in our sensations, but are only inferred from them." (1974: 1)[6] Formulated as a theory of justified belief, we can state this traditional empiricist assumption more exactly as follows:

[6] Quine's attachment to this epistemological picture is distinctive in a number of respects. Quine (1974) emphasizes that he does not see the question of how we are to construct a successful theory of the world from our sensory triggerings as a requirement to vindicate the scientific worldview 'from without'. Instead, he regards it as a question that arises from within scientific practice: in describing the perceptual inputs from which the world is projected, we are free to avail ourselves of information drawn from scientific psychology; the question is whether the inputs, so described, are adequate to the construction of the worldview embodied in science as we know it. The traditional empiricist criterion of justification thus morphs from a foundationalist epistemology grounded in a sense-datum language to a coherentist epistemology designed to ensure the internal coherence of the scientific enterprise.

Classical Empiricism:

Necessarily, for any S, p_i, q: If $\{p_1, p_2, \ldots p_n\}$ is the set of all propositions that S is justified in accepting as reporting the character of her experience and $q \notin \{p_1, p_2, \ldots p_n\}$, S is justified in believing q only if q is analytic or q is inferable from some subset of $\{p_1, p_2, \ldots p_n\}$.

We'll call the conjunction of *Classical Empiricism* with the Quinean view that there are no analytic truths *Radical Classical Empiricism.*

Whatever form of *Classical Empiricism* we accept, explanatory inference will be central to the justification of our beliefs about particulars and properties existing beyond the mind. It just doesn't seem that there are other forms of inference that would allow us to bridge the gap. Plausibly, facts about the way things *seem* don't deductively entail conclusions about the way they *are.* Enumerative induction and singular predictive inference would also appear of little help, as these concern the generality and/or continuation of a previously observed regularity: unless we already know of some correlation between external states and mental episodes, they won't be applicable. Facts about one's mental states could, in principle, support beliefs about external objects if those beliefs were to provide the best available explanation of the character of experience. Just as abductive inference can license belief in subatomic particles to which we have no epistemic access beyond their effects on macroscopic apparatus, so it might license belief in the existence of external objects to which we have no access beyond their effects on the character of subjective experience. Otherwise, we would appear to be out of options.[7]

Note that *Classical Empiricism* doesn't require *all* external posits to earn their rights by featuring in the best explanation of experience. It requires only that they be

[7] For statements of this view see Ayer (1956), Dancy (1985), Goldman (1988), Jackson (1977), Neta (2004), Russell (1912), Vogel (1990).

inferable from this explanation and/or its elements by way of standard canons of reasoning, which needn't themselves be abductive. Thus, it's entirely consistent with *Classical Empiricism* that some of the things we're justified in positing are explanatorily superfluous with respect to our total evidence.

That having been said, it does seem that *Classical Empiricism* implies we cannot be justified in attributing moral qualities to external actions or states of affairs unless moral claims can be vindicated as components of the best available explanation of experience. The argument here is essentially the same as the argument just offered for the view that abduction is needed to license belief in external objects in the first place: other forms of inference don't seem adequate to the justification of any substantive moral conclusions about external states of affairs. To see why this is plausible, let's begin by assuming *Hume's Law*: [8]

Hume's Law:

There is no deductively valid argument whose premises form a consistent set of non-moral propositions and whose conclusion is a non-analytic moral proposition.

On this assumption, if the explanation of experience includes no moral statements, it cannot deductively entail any substantive moral conclusions. In addition, supposing it includes no moral statements, it appears unable to license belief in any moral propositions by way of enumerative induction or singular predictive inference: since these concern the generality and/or continuation of a previously observed regularity, it seems these methods of inference could not support any moral conclusions unless they began from moral premises. Thus, like external objects themselves, it seems that

[8] The statement I offer here is somewhat simplistic and needs to be qualified in a number of ways to avoid certain well-known counterexamples; I've omitted the epicycles to keep us from getting side-tracked. See Kramer (2009: 6-9) for a persuasive defence of a suitably hedged formulation of *Hume's Law* that withstands common objections.

moral properties will have to be vindicated as explanatory posits, rendering them vulnerable, in principle, to the *Anti-Superfluity Principle* if the latter is combined with a value-free explanation of our moral psychology.[9]

For just this reason, *Classical Empiricism* has been a prominent source of moral skepticism in recent meta-ethical debates. In particular, the widely-discussed explanatory problem raised for belief in moral facts by Gilbert Harman (1977), already noted in section 2.1.2 of the previous chapter, is premised on Harman's acceptance of *Radical Classical Empiricism.*[10] According to Harman, a perceptual belief "is reasonable only if it is part of a reasonable conclusion, given your background beliefs along with information about how things look to you." (1973: 173) Since Harman accepts that all observation is theory-laden, he grants that there can be mental episodes which may deserve to be called 'moral observations'. Thus, in *Burning Cat* he grants that you can make the observation that the youths are doing something wrong, just as you would make the observation that they are doing something to a cat: both judgments may arise with the same degree of immediacy, involuntariness, and fittingness with respect to the contents of experience. As noted, Harman says of *Burning Cat* that "an assumption about moral facts would seem to be totally irrelevant to the explanation of your making the judgment you make." (1977: 7). Since he accepts *Radical Classical Empiricism*, he believes that moral observations cannot justify our moral beliefs unless our introspective awareness of these mental states provides a basis for cogent abductive inferences with moral claims as their conclusions. Therefore, according to Harman, the explanatory superfluity of moral facts *vis-à-vis* the phenomena of moral psychology threatens the justification of our moral beliefs.

Notably, Joyce (2006) asks us to consider his evolutionary debunking argument

[9] Here, I am ignoring the following possibility: perhaps some moral claims can be vindicated as part of the description of one's subjective experience, rather than in its explanation. Such moral claims would not be vulnerable to the requirement of inferential justification laid down by *Classical Empiricism.* On introspective ethical knowledge see Sinhababu (ms).
[10] See Harman (1973).

as updating Harman's problem in light of recent evidence about the evolutionary origins of morality.[11] Overall, *Classical Empiricism* seems to provide the required challenge to the otherwise appealing idea that the justification of our moral theories doesn't require that moral values earn their rights as explanatory posits *vis-à-vis* the phenomena of moral psychology. Our next question should therefore be: 'How plausible is *Classical Empiricism*?' In the next subsection, I argue that we should answer: 'Not very.'

2.4 Against Classical Empiricism

While *Classical Empiricism* has been popular throughout the history of modern philosophy, I expect that most philosophers today reject the position. In large part, this is due to the decline of the *Sense-Datum Theory* in the philosophy of perception; historically, this has been the primary motivation for adopting *Classical Empiricism.*[12] I am not going to critique *Classical Empiricism* by attacking its foundations, however. Instead, I want to point to the unacceptability of its implications.[13]

Perhaps the most obvious complaint that could be made against *Classical Empiricism* on this front is that it rules out the possibility of *a priori* justification for non-analytic statements. We might point to a number of apparent counterexamples to this prohibition, such as the *a priori* justified belief that no object can be red all over and green all over at the same time.[14] It has been characteristic of *Ethical Intuitionism* to suppose that moral knowledge rests, at least in part, on *a priori* foundations and may

[11] See also Leiter (2001).
[12] See, *e.g.*, Hume (1748/2007), Locke (1689/1975), Russell (1912).
[13] Jackson (1977) regards the view that the external world must be inferred from the character of experience as "the kind of assumption in terms of which philosophical discussion proceeds rather than to which it is directed. I suspect that assumptions as fundamental as these are finally to be judged not in themselves, but in terms of the edifices erected upon them" (151).
[14] See Van Cleve (1999: 21-27) for a catalogue of non-analytic statements that we appear justified in accepting *a priori*.

be compared at points to our *a priori* knowledge of mathematics.[15] Intuitionists of this sort would naturally reject *Classical Empiricism* as a framework for moral epistemology.

I'm going set aside the issue of the *a priori*, however. There are two reasons. Firstly, the *a priori* is a vexed issue in contemporary epistemology. As Albert Casullo (2003) notes, we seem at present to lack even something as basic as a clear understanding of how to distinguish between perceptual and non-perceptual sources of justification. It's best not to get tangled up in these issues. Secondly, I think it's unclear to what extent allowing for *a priori* knowledge of morality would really evade the problem raised by Harman - and likewise the attempt by Joyce to recast this problem in the form of an evolutionary debunking argument. *Burning Cat* is carefully designed so that one's moral judgment is most naturally described as a perceptual or observational belief. This interpretation has recently been defended by Sarah McGrath (2004). Thus, even if we allow for substantive *a priori* knowledge, we might worry that this won't help in preserving the justification for this particular judgment and others of its kind.

In arguing against *Classical Empiricism*, I'll focus instead on the classic problem of external-world skepticism. If we hope to evade skepticism, *Classical Empiricism* requires that our beliefs about the world beyond the mind can be supported by abductive inferences from premises reporting the character of subjective experience. Common-sense beliefs about the external world certainly offer *an* explanation for the character of our mental life. However, familiar skeptical hypotheses about envatted brains, *malins génies*, *etc.* would seem to do equally well, assuming we have nothing to go on but the introspective reports whose explanation is at issue. As the hypothesis of an external world composed of familiar objects appears to offer no explanatory advantages *vis-à-vis* these skeptical hypotheses, we haven't any justification for

[15] *E.g.*, Clarke (1738/2003), Crisp (2006), Cudworth (1731/1996), Huemer (2005), Ross (1930), Sidgwick (1906/1981).

preferring the former over the latter. Hence, *Classical Empiricism* implies external-world skepticism: we are not justified in believing that there exists a world composed of the familiar objects that appear to us in perception. This, I take it, is an unacceptable result for any epistemological theory, and so we should reject *Classical Empiricism.*

There exists a large literature relating this problem, stretching back over hundreds of years.[16] An exhaustive overview is beyond our scope. Here, I'm going to stick to examining what Harman has had to say about this problem, given his prominence in the debate in which we're engaged. Harman (1973) seeks to evade the skeptical challenge I've just noted by appeal to a form of epistemic conservatism. Although he supposes that external objects must be inferred from internal perceptions, he also believes that we enjoy *prima facie* entitlement to those things in which we already believe. Quite apart from any other considerations, explanations are therefore to be preferred insofar as they fit our pre-existing beliefs. This leads Harman to deny that the skeptic's hypotheses provide equally reasonable explanatory accounts of our perceptions: "The hypotheses the skeptic discusses are not equally reasonable [compared to the posit of an external world], since only one of them is already believed." (1973: 22).

I think Harman is here appealing to an extremely implausible principle of theory choice. To see just how implausible, consider the following real-life case:

Monozygotic Twins:

In November 1999, a woman was raped in Grand Rapids, Michigan. It seemed easy to close the case: DNA was found at the crime scene, and police matched the DNA to Jerome Cooper. There was an unexpected twist. Jerome Cooper had a monozygotic twin, Tyrone. There was otherwise no reason to think either twin more likely to have committed the crime.[17]

[16] See Aune (1991) for an historical overview.
[17] Gilbert (2004).

On Harman's suggestion, it seems the police could still have been justified in continuing to think that Jerome was the culprit: they could say, 'The hypotheses are not equally reasonable, since only Jerome's guilt is already believed.' Obviously, that would be totally unacceptable: they should not have assigned any greater confidence to Jerome's guilt than to Tyrone's. Similarly, if the existence of an external world is otherwise on par with the skeptical hypothesis as an explanation of one's experiences, we can't reasonably prefer the former simply because it is the one we started out believing. Harman's appeal to conservatism fails as a solution to the problem of external-world skepticism.

As I've noted, there's a lot more that could be said about *Classical Empiricism* and external-world skepticism.[18] You could easily write a separate monograph on this issue. I'm going to assume, however, that I've said enough to justify a dismissive attitude toward *Classical Empiricism*. In that case, we should remain unconvinced by the suggestion that our moral beliefs and theories depend for their justification on their value as explanatory posits *vis-à-vis* the phenomena of moral psychology. Since moral theories are not in the business of predicting and explaining these phenomena, we should conclude that *Ockham's Razor* cannot in principle undercut their justification if combined with value-free evolutionary accounts by which these phenomena *are* explained. Ethics is not psychology. For a successful debunking argument we must look elsewhere.

3. Insensitivity

The line of argument we considered in the previous section turned on the assumption that the justification of our moral beliefs depends positively on their inclusion in the

[18] See in particular the sophisticated defence mounted by Vogel (1990).

best available explanation of our moral psychology. An alternative approach would be to appeal to some form of negative dependence: we might think the justification of our moral beliefs depends on our having *no reason to deny* that moral facts figure in their explanation. Evolutionary explanations might be thought to provide defeaters by offering reasons to deny just this.

3.1 Insensitivity and evolutionary debunking arguments

One way to develop this suggestion would be by appeal to considerations of Nozickian sensitivity. Following Nozick (1981), we say that *S*'s belief that *p* is *sensitive* if were *p* false, *S* would not believe that *p* and *insensitive* otherwise. We might think that even if our moral beliefs happen to be true, the fact that these beliefs are explained in evolutionary terms without reference to the existence of any corresponding moral facts implies we would have these beliefs even if they were false: they would still have been selectively advantageous, after all. Nozick proposes that *S* knows that *p* only if *S*'s belief that *p* is sensitive.[19] We might feel inclined to accept the following principle:

Insensitivity Defeat:[20]

Necessarily, for any *S*, *p*: If *S* believes *p* with *prima facie* justification but has sufficient reason to believe that her belief in *p* is insensitive, this constitutes a defeater for her continued acceptance of *p*.

Insensitivity Defeat sounds very attractive on the face of it. If my belief in *p* is

[19] Here I am neglecting certain epicycles, such as the relativization to methods (Nozick 1981: 179-185). I take up this issue in the next subsection.
[20] Cf. Pollock (1986).

insensitive, it would seem that I'm not really able to tell whether *p* is true or false;[21] and if I ought to think that I'm not able to tell whether *p* is true or false, it seems plausible that I shouldn't trust my judgment with respect to *p*. *Insensitivity Defeat* also appears to provide a very natural and satisfying explanation of what's going on in cases of undercutting defeat such as the following:

Parental Bias:

You believe that your daughter is a talented pianist. However, you learn that, as a parent, you are biased in favour of holding very positive views about the abilities of your children.

On the face of it, this information is defeating because (or insofar as) it implies you would believe that your daughter is a talented pianist even if she were mediocre or just plain bad.

If *Insensitivity Defeat* is true and evolutionary explanations provide sufficient reason to suppose that our evolved moral beliefs are insensitive, we should then conclude that evolutionary explanations are (to that extent) debunking. This section considers whether we ought to accept the first conjunct in the antecedent of this conditional. In the next section, I will argue that we should in any case reject the second.

3.2 Some examples

A number of philosophers go in for sensitivity-based debunking arguments. For example, Parfit (2011: 525-526) maintains that from the premise

> These beliefs would have been advantageous whether or not they were true

[21] Cf. Becker (2012): "sensitivity captures an important feature of knowledge, namely the ability to tell the difference between when a proposition is true and when it is not." (82)

we may infer that

> Natural selection would have disposed us to have these beliefs whether or not they were true

from which we should conclude

> We cannot justifiably believe that these beliefs are true.

Similar reasoning has been offered by Joyce and Ruse. Ruse (1986) tells me:

> You would believe what you do about right and wrong, irrespective of whether or not a 'true' right and wrong existed! … Given two worlds, identical except that one has an objective morality and the other does not, the humans therein would think and act in exactly the same ways. (254)

Similarly, Joyce (2001) writes:

> Suppose that the actual world contains real categorical requirements – the kind that would be necessary to render moral discourse true. In such a world humans will be disposed to make moral judgments …, for natural selection will make it so. Now imagine instead that the actual world contains no such requirements at all – nothing to make moral discourse true. In such a world humans will *still* be disposed to make these judgments …, just as they did in the first world, for natural selection will make it so. What this shows is that the process that generates moral judgments exhibits an independence relation between judgment and truth, and these judgments are thus unjustified. (163)

None of these authors formulates any explicit principle of defeat to undergird his argument, but it seems plausible that each relies implicitly on something like *Insensitivity Defeat*. It may be that all three take *Insensitivity Defeat* to be so obviously true as not to be worth stating. However, the principle is subject to a number of epistemological difficulties, as I explore in sections 3.4 and 3.5. Before that, I want quickly to go over an issue relating to the semantics of subjunctive conditionals.

3.3 Subjunctive conditionals and (im)possible worlds

Claims about the sensitivity or insensitivity of our beliefs, moral or otherwise, involve subjunctive conditionals. Ignoring certain minor disagreements between Stalnaker (1968) and Lewis (1973), the most popular semantics for subjunctive conditionals is as follows:

Lewis-Stalnaker Semantics for Subjunctive Conditionals:

For any *p*, *q*: *If p were the case, then q would be the case* is true iff if *w* is the closest possible world to the actual world in which *p* is true, *q* is true at *w*.

On this kind of semantics, it will be difficult to see how evolutionary explanations could show our moral beliefs to be insensitive, at least provided we have antecedent justification for supposing the moral propositions we believe are true and that moral facts supervene on non-moral facts.[22] We can state the latter assumption as follows:

Moral Supervenience:

For any possible worlds *w*, *w'*: *w* and *w'* differ in terms of their moral facts only if *w* and *w'* differ in terms of their non-moral facts.

Granting these assumptions, we can reason as follows. Suppose the actual world is one in which Muammar Gaddafi is an evil man, just as we believe him to be. Assuming *Moral Supervenience*, we should think a possible world in which Gaddafi is not an evil man must differ from the actual world in certain relevant respects. For example, it might be one in which he devotes his life to improving child literacy and never

[22] Wielenberg (2010) raises this issue. Cf. Sturgeon (1985) on moral explanations and supervenience.

becomes the dictator of Libya. In close possible worlds of that kind, it seems implausible that any of us believes that Gaddafi is an evil man. Furthermore, evolutionary accounts of the origins of our moral beliefs provide no reason to think otherwise, even if they explain why we are disposed to judge men like Gaddafi to be evil: such accounts provide no reason to think that we would judge a man who devotes his life to improving child literacy as morally reprehensible. How, then, could they support a verdict of insensitivity?

It's plain that what Joyce and Ruse have in mind is something like the following. They think that evolutionary explanations of our moral beliefs support the conclusion that moral facts play no role in bringing about our moral beliefs. Thus, if we stripped off whatever moral facts there are in the actual world and otherwise left everything that's independent of the moral facts just the same, we would have human beings with the very same moral beliefs, albeit false beliefs. In this way, they suppose, evolutionary explanations support the claim that certain of our moral beliefs are insensitive: it's not true that we would not hold these beliefs were they false. However, assuming *Moral Supervenience*, a world in which we strip off whatever moral facts there are *and leave everything otherwise the same* is *not a possible world*. It's therefore not relevant to assessing the truth of the relevant subjunctive conditional, assuming the Lewis-Stalnaker framework.

As this suggests, the claim that our evolved moral beliefs are insensitive won't go through unless we reject or modify the Lewis-Stalnaker semantics. This might not worry some, as it's already well-known that this semantics has peculiar implications for subjunctive conditionals whose antecedents are false in all possible worlds: it implies that all such *counterpossibles* are trivially true, including *If Russell had axiomatised mathematics, all credit would have gone to Nat King Cole*. Those who are inclined to reject the view that all counterpossibles are trivially true might broaden the

Lewis-Stalnaker framework by simply including *impossible worlds* within it and supposing that *If p were the case, then q would be the case* is true iff if *w* is the closest possible or impossible world in which *p* is true, then *q* is true at *w*. This could eliminate the problem I've just noted for Joyce and Ruse, as Joyce (forthcoming) suggests. Frameworks of this kind have been proposed by Daniel Nolan (1998), Graham Priest (1997), and Takashi Yagisawa (1988, 2010), but are subject to heavy criticism from Tim Williamson (2007: 171-175).

Sadly, this controversy is beyond my expertise. For the sake of argument, I'm happy to go along with the assumption that the correct semantics for subjunctive conditionals poses no obstacle to the view that our evolved moral beliefs are insensitive. Let's turn instead to some pertinent epistemological worries about sensitivity-based debunking arguments.

3.4 Troubles with skepticism

In the previous section, I argued that we ought to reject *Classical Empiricism* because it leads to external-world skepticism. The same complaint has been lodged against *Insensitivity Defeat*. Roger White (2010) says: "the principle is almost certainly false as it stands. For it quickly leads to the conclusion that I'm not justified in believing much of anything." (581) As Nozick (1981: 200-204) made clear, the denials of skeptical hypotheses are one and all insensitive. Consider the familiar *BIV Hypothesis*: I am just a brain in a vat (BIV) being stimulated so as to experience a seamless hallucination as of sitting at my computer, typing these words. I believe the *BIV Hypothesis* is false, but I would believe this even if it were true. What is more, I'm well aware of this. Thus, by *Insensitivity Defeat*, I'm not justified in denying the *BIV Hypothesis*; I'm not justified in believing that I'm not merely a BIV being stimulated by clever neuroscientists so as to hallucinate the objects I see before me.

It doesn't follow straightforwardly from this that I'm not justified in believing much of anything. Just as Nozick (1981) and Dretske (1970, 2005) deny that knowledge is closed under competent deduction, someone who accepts *Insensitivity Defeat* could insist that closure fails for epistemic justification.[23] In this way, they could maintain that I'm justified in believing that I have hands, though I'm not justified in denying the *BIV Hypothesis*, despite the fact that I can deduce the latter from the former. They might even appeal to the plausibility of *Insensitivity Defeat* in order to motivate closure denial. Here's how this might go. Although the denial of the *BIV Hypothesis* is insensitive, I clearly have no reason to suppose that my belief that I have hands is similarly insensitive. Since there is no corresponding defeater for the belief that I have hands, we should think this belief remains justified in spite of the fact that the denial of the *BIV Hypothesis* is defeated by my awareness that its denial would be insensitive. Thus, I can believe the former but not the latter, though the former entails the latter.

Closure denial may be too much for some, just as many philosophers cannot stomach it when applied to knowledge.[24] On the approach I've just sketched, it seems we ought also to think that I'm justified in believing the conjunction *I have hands and I am not a BIV*, even though I'm not justified in believing *I am not a BIV*. *Insensitivity Defeat* implies no defeater for belief in the conjunction. The closest possible world in which the conjunction is false will be a world in which I lack hands, but we have no reason to suppose I would still believe the conjunction in that world: presumably, I would reject the conjunction because I would reject the claim that I have hands. Thus, even though I can't believe *I am not a BIV*, it seems there is no reason to think I can't justifiably believe *I have hands and I am not a BIV*. The conclusion that I can believe this conjunction without being justified in accepting one of its conjuncts might strike us as

[23] See Schechter (2013) for an independent argument against closure for justification.
[24] *E.g.*, deRose (1995), Hawthorne (2004, 2005).

especially outrageous.

There is another – and, in my view, more plausible – means by which those attracted to *Insensitivity Defeat* could try to allay their troubles with skepticism. Although Nozick initially proposes simply that knowledge requires sensitivity, he quickly qualifies this position in order to accommodate the following well-known counterexample:

The Sick Grandchild:

A grandmother visits her grandson in hospital and sees that he is well. If he were dead or sick, others would tell her he was well in order to spare her grief, and she would believe them.

Nozick wants to say, of course, that the grandmother knows that her grandson is well, though this belief is in fact insensitive. To deal with this case, he modifies the account by introducing a reference to the method of belief-formation. On this new proposal, *S* knows that *p* via method *M* only if were *p* false and *S* used *M* to decide whether or not *p*, *S* would not believe that *p* via *M*. Thus, the grandmother counts as knowing her grandson is well although her belief is insensitive, because she would not believe that her grandson is well were she to visually inspect him and he was sick or dead. As an alternative to Nozick's modification, we might adopt the following suggestion, due to Tim Williamson (2000: 154): *S* knows that *p* via method *M* only if were *p* false, *S* would not believe that *p* via *M*.

How does this help with the issue of skepticism? As Williamson points out, it's much less straightforward to verify that the denials of skeptical hypotheses fail a sensitivity-based criterion that includes some form of methods-relativization. Everything depends on how methods are individuated. Is the method by which I form the belief *I am not a BIV* the same in both cases - the bad case in which the *BIV*

Hypothesis is true, and the good case in which it is false? Suppose that methods are individuated 'internally', such that two individuals count as using the same belief-forming methods if they are alike in terms of their (non-factive) mental states. Then, the answer will be 'Yes'. However, if methods are individuated 'externally', the answer may well be 'No'.

Thus, in order to avoid the skeptical troubles that seem to arise in light of *Insensitivity Defeat*, we might adopt the following analogous modification:

Methods-Relative Insensitivity Defeat:

Necessarily, for any *S*, *p*: If *S* believes *p* with *prima facie* justification but has sufficient reason to believe that she would believe *p* via the same method even if were *p* false, this constitutes a defeater for her continued acceptance of *p*.

We could then couple this principle with an externalist account of methods-individuation under which the good case and the bad case don't involve the same method of belief-formation, and the problem of skepticism would be sidestepped. This requires, of course, that we have some independent motivation for individuating methods in this way.[25] Philosophers who are otherwise opposed to externalist approaches in epistemology might well suppose there are none, or none that are good enough. I don't want to dwell on this issue, however, as there are further epistemological problems for *Insensitivity Defeat* that can't be avoided by switching to *Methods-Relative Insensitivity Defeat*.

3.5 Further epistemological difficulties

Both *Insensitivity Defeat* and *Methods-Relative Insensitivity Defeat* appear to suffer from

[25] Williamson (2000) provides some motivation on this front with respect to sensitivity and knowledge.

difficulties associated with inductive inference - problems analogous to those raised by Jonathan Vogel (1987) in attacking Nozick. Consider the following example due to Vogel:

Melting Ice:

You leave some ice out in your back-garden on a hot summer's day. An hour later, you come to believe the ice has melted. If it hadn't melted (*e.g.*, because the heat of sun had somehow failed to find its way to the ice), you would still believe that it had melted, and on the very same grounds.

Vogel suggests that this is a case in which you know that the ice has melted. Others agree about this and similar cases.[26] We probably feel even more confident that you are *justified in believing* that the ice has melted, even if you know you would still believe this even if the sun had somehow failed to find its way to the ice. *Insensitivity Defeat* and *Methods-Relative Insensitivity Defeat* imply the opposite: you are not justified in believing that the ice has melted.

Those who defend a sensitivity condition on knowledge have replied to cases like *Melting Ice* by insisting that although the condition rules out knowledge that the ice *has in fact* melted, it is compatible with the knowledge that the ice has *probably* melted, and so the counterexample is not that damaging overall.[27] In the same vein, it could be said that although you are not justified in believing that the ice has melted, according to our favoured sensitivity-based condition of defeat, you are justified in believing that the ice has *probably* melted, which goes a long way to salvaging our intuitions about the case.

There is a hitch to this reply. It's natural to read the belief that the ice has

[26] *E.g.*, Hawthorne (2004), Sosa (1999).
[27] Becker (2007); Roush (2007).

probably melted as to do with objective probability, rather than subjective or epistemic probability: the belief is about the chance of the event, rather than what confidence we assign or should assign to it. Events which have already occurred or failed to occur should be assigned objective probability 1 or 0.[28] Thus, if it is now the case that the ice has either melted or failed to melt and you're not justified in believing that the ice has *in fact* melted, you can't believe that the objective probability of the ice having melted is currently high. At best, you can believe that there *was* a high objective probability for the ice to melt. This would seem to do even less in terms of salvaging our intuitions.

This is a bad result, but I can find some sympathy for those who want to retain a sensitivity-condition for defeat. As I noted in 3.1, a principle of this kind is attractive in certain respects. The problems we've encountered here aren't obviously sufficient to outweigh these benefits. At least, I can imagine reasonable people taking this view. I can also find sympathy for those who would rather reject *Insensitivity Defeat* and *Methods-Relative Insensitivity Defeat.*

For those who are in the first camp, I want now to offer a non-epistemological objection to sensitivity-based debunking arguments. These arguments, I claim, rely on the *Total/Functional Fallacy.* The same is true of arguments relying on *Ockham's Razor.* The next section is devoted to exposing this fallacy and the role it plays in contemporary evolutionary debunking arguments.

4. The *Total/Functional Fallacy*

Proponents of sensitivity-based debunking arguments assume, as I've noted, that if we strip off whatever moral facts there are and otherwise leave everything that is explanatorily independent of the moral facts the same, we would have human beings

[28] Eagle (2012).

with the very same moral beliefs, formed via the very same methods. Similarly, those who appeal to *Ockham's Razor* assume, in Joyce's (2006) words, that "we have a complete explanation of moral judgment with no need to posit any extra ontology in the form of moral facts." (188) However, conclusions of this kind cannot be inferred from *Functional Truth-Irrelevance*, nor from any other account of our moral beliefs in terms of their ultimate, evolutionary causes.

Here is a neat hypothetical example to show that an explanatory factor might be irrelevant for explaining why a certain trait was selectively advantageous, while being crucial to the explanation of the trait's existence:

Giol's Spearpoint:

The island of Niron is divided among a number of competing, small-scale societies. The spear-maker of the Orin people, Giol, modifies the design of a spearpoint because he finds the resultant shape beautiful. So modified, the spear-tip happens to be unusually deadly. The Orin people conquer their neighbours, establishing colonies to occupy their land. After ten generations, everyone in Niron is a descendant of the original Orin culture, and everyone uses the spearpoint. Craftsmen continue to manufacture spearpoints of this design based on the feeling that it is uniquely beautiful.

The selective advantage conferred by the spearpoint in this example is independent of the aesthetic preference of any craftsman and entirely to do with increased lethality. Nonetheless, the aesthetic preferences of the craftsmen are a significant part of the explanation for why the Orin people make their spearpoints according to this design. As explanatory factors, selection and aesthetic preference are entirely complementary. Furthermore, supposing some anthropologist knew why the Orin culture and its spear design had been favoured by selection, she would not be justified in concluding that the aesthetic preferences of Orin craftsmen played no role in explaining the production of

spearpoints of this design. No such conclusion follows.

If this holds in *Giol's Spearpoint*, there is no reason to expect things to be otherwise when it comes to natural selection and human morality. Grant that where a disposition to adopt certain moral beliefs has been favoured by selection, the truth of these beliefs will be irrelevant to the explanation of their having been so favoured. This means that moral facts play no part in the functional explanation of these beliefs or whatever psychological structures bias their acquisition. It doesn't follow that moral facts can't play a role in explaining the very same moral beliefs in terms of more proximate factors. Perhaps individual men and women formed (and continue to form) such moral beliefs as were selected for by responding to the rightness or wrongness of various actions. Nozick (1981) appears to make just this suggestion:

> If ethical behavior increases inclusive fitness, this will explain the spread of such behavior in the population. Yet each individual's behavior, ancestor or descendant, might be explained by her recognizing certain ethical truths and acting on them. What spreads in the population can be the capacity to recognize certain (ethical) truths and the predisposition to act on this recognition. (345)[29]

Many philosophers working on evolutionary debunking arguments seem to ignore this possibility. For example, Parfit (2011) blatantly conflates the distinction between proximate and ultimate causes, asking us to consider "which aims and acts would have been reproductively advantageous, so that it might be natural selection, and not our response to reasons, that motivated early humans to act in these ways" (526-7). Similarly, some philosophers move immediately from the claim that moral facts aren't needed to explain why some element of our moral psychology was selectively advantageous to the conclusion that moral facts are irrelevant in accounting for any corresponding moral beliefs.[30] Here is an example. Walter Sinnott-Armstrong (2006a)

[29] Cf. Johnston (2001), McMahan (2002: 166-168).
[30] Joyce (2006) does pay some attention to the issue of proximate causes. He writes: "I have been

considers a number of evolutionary explanations for common moral beliefs. He discusses the hypothesis that human beings who believe they are morally required to help others will more reliably help and so will reap greater rewards from relations of mutual reciprocity. He offers similar explanations for beliefs that prohibit killing, lying, and promise-breaking, writing: "Humans with such moral beliefs will be less likely to kill, lie, etc. Thus, given the background tendency to reciprocity, they will be less likely to be killed, deceived, etc. so they will be more likely to survive and reproduce." (42) Having set out these proposals, he writes: "If anything like this evolutionary story does explain common moral beliefs, then there is no need to postulate moral facts to explain those beliefs." (43) This is not correct. At best, it follows that moral facts are not needed to explain *why these beliefs were selectively advantageous*. Sinnott-Armstrong commits the *Total/Functional Fallacy*: he infers that since moral facts are irrelevant to the explanation of some feature of our moral psychology in terms of its evolutionary function, they are irrelevant *tout court* in explaining any associated moral beliefs.

He is not alone in this. The very same fallacy appears to be at work in all of the debunking arguments we've considered in this chapter. They all assume that evolutionary explanations can establish that moral facts do not figure at all in the explanation of some subset of our moral beliefs. To the extent that evolutionary explanations are confined to outlining the ultimate causes of our moral beliefs, this would seem impossible. Such explanations leave open the question of whether moral facts might explain our beliefs at the level of their proximate causes.

Philosophers may be convinced on independent grounds that this question

---

addressing a 'why' question: Why would natural selection come up with the trait of moral thinking? But even a complete answer to this … would leave us with a further question: *How* would natural selection bring about moral thinking?" (123) In addressing this question, however, Joyce considers only the internal side of things. In light of evidence for the importance of emotion in moral judgment, he maintains that we have "a very coarse-grained answer to what natural selection did to the human brain to enable moral judgment: It manipulated emotional centers." (125) Whether emotions figure as the internal proximate causes of moral judgments, this fails to address the question of whether instantiations of moral properties sometimes act as external proximate causes of our moral beliefs.

should be answered one way or the other. As I've noted, the question raised by Harman about whether your judgment in *Burning Cat* is explained by the fact that what the children are doing is wrong is really a question about whether moral facts are relevant to explaining this moral verdict in terms of its proximate causes. In the ensuing discussion, a number of philosophers – both realist and anti-realist – concluded that moral facts were explanatorily irrelevant in just the way Harman claimed, and others adopted the contrary view. Those who are in the first camp might not be so worried by the problem I've noted. Having already decided that moral facts are explanatorily irrelevant at the level of proximate causes, they might simply treat evolutionary accounts as closing off the possibility that these same facts play a role when it comes to the ultimate causes of our moral beliefs. I doubt very much that they were especially worried about this possibility in any case. As Sturgeon (2006) notes, it's an implicit assumption of Harman's discussion that his examples provide a kind of best-case scenario: "if he is right about his examples, then his conclusion must apply more generally, to all moral beliefs: unlike non-moral beliefs, our holding them is never to be explained, even in part, by assuming them to be true." (242) In any case, those who were not already convinced by Harman's view should be unimpressed by the arguments we've been considering. The same holds for those who have yet to make their minds up. To the extent that we have sufficient reason to conclude that our evolved moral beliefs are not explained (partly, at some level) by moral facts, these reasons must go beyond what is available in light of functional accounts of the origins of morality. Thus, both of the arguments that we have been considering in this chapter would appear to rely on mistakes regarding the limitations of natural selection explanations.

5. Conclusion

In this chapter, I've argued against two lines of reasoning purporting to show that evolutionary accounts of morality are debunking. The first appealed to *Ockham's Razor* - more exactly, to the *Anti-Superfluity Principle*. The second relied on considerations of Nozickian sensitivity; *Methods-Relative Insensitivity Defeat* emerged as the most plausible basis for an argument of this kind. In sections 2 and 3, I considered a range of epistemological problems. I noted that our moral beliefs don't appear vulnerable to the *Anti-Superfluity Principle* unless our moral views depend for their justification on their value as explanatory posits *vis-à-vis* the phenomena of moral psychology. While *Classical Empiricism* supports a view of this kind, I argued that *Classical Empiricism* should be rejected in light of its difficulties with external-world skepticism. Moving on to the second line of argument, the principle I called *Insensitivity Defeat* also turned out to generate skeptical problems. I noted that those attracted to sensitivity-based defeat-conditions could potentially evade these problems by switching to *Methods-Relative Insensitivity Defeat*. This switch didn't offer any advantage in terms of evading the problems with inductive inference that I noted in 3.5. In section 4, I explained that both lines of argument should be rejected in any case. Both assume that we can rely on evolutionary accounts to show that moral facts do not figure in the explanation of our moral beliefs. This assumption, I suggested, is driven by the *Total/Functional Fallacy*. More generally, I noted that evolutionary accounts appear to leave open the question of whether moral facts figure in the explanation of our moral beliefs at the level of proximate causes – the question raised by Harman. Philosophers who have thought that we can rely on evolutionary theories to answer Harman's question seem to have confused the distinction between proximate and ultimate causes.

I conclude that we can safely ignore evolutionary debunking arguments appealing to *Ockham's Razor* and Nozickian sensitivity: they are epistemologically dubious and appear to rely on confusions about the nature of biological explanation.

These arguments should not convince us to give up any of our moral beliefs.

4.

The Coincidence Problem: a skeptical appraisal

1. Introduction

In the previous chapter, I rejected two lines of argument, each purporting to derive debunking implications from *Functional Truth-Irrelevance*. I noted some purely

epistemological problems for these arguments, but also pointed to a graver, shared flaw. The arguments assumed that since we can explain why (psychological structures dispositive of) certain moral beliefs/intuitions would have been selectively advantageous regardless of their truth or falsity, moral facts are explanatorily irrelevant in accounting for the relevant beliefs/intuitions. This inference, I noted, relies on a confusion of proximate and ultimate explanatory factors.

Some philosophers have claimed that evolutionary debunking arguments must generally rely on the claim that evolutionary explanations show moral facts to be explanatorily irrelevant altogether in accounting for some subset of our moral beliefs.[1] If this were true, my comments in the last chapter would suffice to show that no such argument should convince us to give up any of our moral beliefs. However, not all debunking arguments rely on this premise. The argument from *Phyletic Contingency* that I develop in chapters 5 and 6 doesn't. Nor does the argument – or family of arguments – to be discussed in this chapter.

The family of arguments we're going to consider here are centred on what I call the *Coincidence Problem*. Roughly speaking, the problem is as follows. It may seem to require an extraordinary coincidence if natural selection has favoured the evolution of moral beliefs which turn out also to be true, given that matters of truth and falsity are irrelevant in accounting for the selection-pressures shaping human moral psychology. Since we cannot reasonably expect such a coincidence (we might think), we ought to suspend judgment with respect to those beliefs. As this statement suggests, the problem is narrowly focused on issues to do with the causes of selection and their explanatory relation (or lack thereof) with respect to the ethical facts. It does not turn on issues to do with proximate causes.

[1] *E.g.*, Mason (2010), Wielenberg (2010).

This *Coincidence Problem* is most prominent in a series of articles by Sharon Street (2006, 2008a, 2011), though Street herself believes it arises only if we assume some form of *Meta-Ethical Realism*; Street herself endorses *Meta-Ethical Constructivism*, according to which ethical facts are constitutively dependent on our attitudes. The problem also features in work by other philosophers hoping to derive rather different conclusions from Street, as we'll see in the next section. I think the *Coincidence Problem* is by far the most promising basis for a debunking argument grounded in *Functional Truth-Irrelevance*. Nonetheless, the *Coincidence Problem* is ultimately illusory. Although a coincidence may be required, this is unproblematic and provides no reason to revise our beliefs. It's the aim of this chapter to substantiate this appraisal.

In the next section, I'll explore recent work by three authors - beginning with Street - each of which embodies the *Coincidence Problem* in some form. The rest of the chapter is devoted to undoing the *Coincidence Problem*. In section 3, I examine more carefully what sort of correlation between ethical truth and selective advantage we may have to posit, but supposedly cannot - at least not if we are realists. In section 4, I examine whether we are in fact forced to posit some form of coincidence in order to maintain that such a correlation exists, and whether our preference for *Meta-Ethical Realism* over *Meta-Ethical Constructivism* could make a difference on this point. The answers are 'yes' and 'no', respectively. In sections 5 through 8, I consider whether it is really unreasonable to posit a coincidence of this sort; I argue that it is not. Central to my argument will be analogy between the *Coincidence Problem* and the *Fine-Tuning Problem* in cosmology. This analogy will serve to illuminate both strengths and weaknesses of the *Coincidence Problem*. The weaknesses, I claim, far outweigh the strengths.

2. Locating the problem

My aim in this section is to explore recent work by three philosophers to help isolate the *Coincidence Problem* as a feature of contemporary debate on evolutionary debunking arguments and show how the problem can be applied in arguing for quite disparate conclusions in meta-ethics and/or moral epistemology.

2.1 Some examples

2.1.1 Street

I'll begin with Sharon Street's work. As debunking arguments go, hers is atypical in two respects. Firstly, though most philosophers working in this area are principally interested in debunking arguments targeting our moral beliefs, Street's concern is with beliefs about practical reasons in general. To take account of this more expansive target, I'll here speak in terms of 'ethical beliefs' and 'ethical facts' understood according to the broadened definition suggested by Bernard Williams (1986) and Roger Crisp (2006), taking these to encompass beliefs and facts about practical normativity, including, but not limited to, morality. Street believes that our ethical beliefs/intuitions are influenced to a great extent by certain innate biases: so much so that rejecting those of our beliefs which have been influenced in this way would be catastrophic from the perspective of our having any reasonable grip on how we ought to live. The second notable feature of her argument is that she thinks realists - and *only* realists - find difficulty in accepting evolutionary accounts of our ethical beliefs without being forced into this kind of catastrophic normative skepticism. Street believes, as noted, that normative facts are constructed out of our attitudes: facts about what a person ought to do are constituted by facts about the reasons she would self-ascribe when her normative judgments are brought into reflective equilibrium. On this picture, value arises out of our valuing. Street believes anyone accepting a view of this

kind faces no difficulty in accommodating evolutionary explanations of our ethical beliefs.

With these preliminary points in mind, let's now set out her argument. We start off from the assumption that "the forces of natural selection have had a tremendous influence on the content of human evaluative judgements." (Street 2006: 113) Given this assumption, Street confronts the realist with a 'Darwinian dilemma': "either the evolutionary influence tended to push our normative judgments *toward* the independent normative truth, or else it tended to push them *away from* or *in ways that bear no relation to* that truth." (2011: 12) In other words, of the normative beliefs favoured by natural selection, either these have tended to be true, or else they have not.

Given the assumption that our fundamental normative principles have been determined by natural selection, the denial of such a tendency may seem clearly problematic: our normative outlook would be saturated with error. This might not be so bad if we had some independent means of weeding out these errors, but Street (2006: 124) denies the existence of any such capacity. Rational reflection, she claims, is limited to systematizing the evaluative judgments resulting from our evolutionary inheritance.

Suppose, then, that the realist takes the second horn: she posits that there *has* been a consistent correlation between those affective dispositions which have been fitness-maximizing and those which are conducive to believing mind-independent normative truths. Then, Street insists, the realist has some explaining to do. "This degree of overlap between the content of evaluative truth and the content of the judgements that natural selection pushed us in the direction of making begs for an explanation." (2006: 125) This explanatory burden, Street suggests, cannot be met: "the realist about normativity owes us an explanation of this striking fact, but has none" (2008a: 207).

Street maintains that the only explanatory strategy open to the realist is to adopt the *Tracking Account*, which I'll interpret here as the view that there has been selection for (psychological structures dispositive of) true normative beliefs.[2] Street maintains that, for the realist, "the tracking account is the only (non-coincidence-positing) way of seeing how evolutionary forces could have pushed our values *toward* independent normative truths" (2011: 13). The tracking account is opposed to what she calls the *Adaptive Link Account*, according to which matters of truth and falsity are irrelevant in explaining the fitness-benefits associated with these biases: such adaptive advantages as they confer are explained entirely by their motivational effects and by the independent fitness-benefits associated with the behaviours they motivate. Street rejects the *Tracking Account*: it is, she says, "scientifically indefensible." (2011: 13)

Street concludes that realists cannot grasp the second horn of the dilemma, having no means to explain the existence of a consistent correlation between those affective dispositions which have been fitness-maximizing and those which have been conducive to believing normative truths. Anti-realists, she argues, face no similar difficulty, since they count the normative facts as a function of our attitudes, and our attitudes are a function of our evolutionary history:

> Antirealism explains the overlap not with any scientific hypothesis such as the tracking account, but rather with the metaethical hypothesis that value is something that arises as a function of the evaluative attitudes of valuing creatures – attitudes the content of which happened to be shaped by natural selection. (Street 2006: 154)

2.1.2 Rosenberg

[2] As the reader will recall from chapter 2, Street actually wavers in her formulation of the *Tracking Account*, fudging the selection-for/of distinction. The interpretation I've adopted here is, I believe, the most appropriate for the issues at hand.

Alex Rosenberg (2011) relies on the *Coincidence Problem* to argue for a rather different conclusion to Street: that we should give up those moral norms which are known to have evolved under the influence of natural selection.

Rosenberg begins from the assumption that certain core norms have evolved because they raised the relative fitness of their subscribers. We, as their descendants, suppose these norms are correct: they capture what we are in fact morally required to do. This raises the crucial question: "Is the correctness of core morality and its fitness a coincidence?" (Rosenberg 2011: 109) One possibility is that there is no coincidence because there is an explanatory link running from the former to the latter: these norms were favoured by natural selection because they correspond to the moral truth. Rosenberg rejects this view as scientifically indefensible. Another possibility is that the relevant norms are true because they increased the relative fitness of their subscribers. As Rosenberg notes, this hardly seems plausible: "There doesn't seem to be anything in itself morally right about having lots of kids, or grandchildren, or great grandchildren, or even doing things that make having kids more likely. But this is all the evolutionary fitness of anything comes to." (110) Rosenberg concludes that we "cannot explain the fact that when it comes to the moral core, fitness and correctness seem to go together." (113) Since we can't reasonably believe that the two just happen to overlap, we have to give up viewing the moral norms in question as correct. Because the relevant principles are so central to our moral outlook, Rosenberg believes that a general moral skepticism results from their abandonment.

2.1.3 Huemer

As a final example of the *Coincidence Problem*, we'll consider a recent paper by Michael Huemer (2008). Huemer is an ethical intuitionist: he accepts a non-naturalist, realist meta-ethics and believes that moral intuitions confer (defeasible) justification on beliefs

of the same content.[3] However, he rejects the conservatism typically associated with intuitionism. Intuitionists, he maintains, should adopt "revisionary ethical views, rejecting a wide range of commonly accepted, prephilosophical moral beliefs." (369) In many cases, he believes, the defeasible justification conferred by our intuitions can be soundly defeated. Among those factors that should lead us to discount our intuitions, he counts the influence of 'biological programming'. In support of this view, he writes:

> an organism's reproductive fitness would seem to be best promoted by its having values skewed in a certain direction: by the organism's taking its own reproductive success, or things normally correlated with one's own reproductive success, to be good, whether or not those things are objectively good. … For this reason, it would seem that if the values toward which natural selection biased us coincided with the objectively correct values, this would be sheer coincidence. Such a coincidence cannot reasonably be expected. (377)

Unlike Street or Rosenberg, Huemer believes that the revisions in our moral outlook required by this conclusion are not so wide-ranging as to yield full-blown ethical skepticism. The result is austerity, not collapse: "a shift that leaves us with some moral beliefs, but perhaps with a very different set from those that more traditional methods lead to." (379) Huemer suggests that his 'revisionary intuitionism' should incline us to accept some form of consequentialism, as consequentialist theories are likely to be driven by abstract, formal intuitions, not readily traceable to biasing factors like biological programming.

2.2 The problem distilled

There are, as stated, noticeable differences among the authors discussed, but also a recognisable common thread. All accept *Functional Truth-Irrelevance*: they maintain that those ethical beliefs which have been favoured by natural selection have been

[3] See Huemer (2005).

favoured without regard for their truth or falsity. Nonetheless, we think, the resultant beliefs *are* true. And here lies the problem. We seem committed to thinking that natural selection just happened, coincidentally, to favour the evolution of the right ethical judgments. And this may sound like too much to be believed. This is the *Coincidence Problem.*

As I see it, there are three possible responses to this problem. Firstly, we could try to find a solution that avoids the commitment to a merely coincidental overlap but allows us to maintain belief in *Functional Truth-Irrelevance* and in the truth of our evolved ethical beliefs. Street takes this route. She maintains that the need to posit a coincidental overlap can be avoided by rejecting a realist account of value.[4] A second possible response is to concede that we ought to make substantial revisions in our normative outlook, rejecting any of our ethical beliefs known to be explained by natural selection. This is the line taken by Rosenberg and Huemer. There is disagreement as to the scope of the revisions required under this second option: Huemer maintains that the upshot is something less than ethical skepticism; Street and Rosenberg disagree. A third possible response is the one I favour: deny the existence of a problem and say that we can happily live with a coincidence. The suggestion that we cannot reasonably believe that natural selection just happened to favour the evolution of objectively correct ethical beliefs may be intuitively compelling, but we shouldn't rest content with our intuitions on the matter. It is well-established that our intuitions about coincidences are unreliable: we are far too quick to think that meaningful correspondences can't be merely accidental.[5] We should consider more carefully whether our intuitions stand up to scrutiny. I argue that they do not.

[4] Huemer says that "if the values toward which natural selection biased us coincided with the *objectively* correct values, this would be sheer coincidence" (my emphasis), suggesting that he too sees the rejection of objectivism as one possible route around the problem, albeit not the one he prefers.
[5] Kahneman (2011: 114-118); Zusne & Jones (1989).

3. The initial fund and the derived fund

Unless we reject *Functional Truth-Irrelevance* or revise our ethical beliefs, we seem committed to an implausible coincidence, at least if we are realists. This is the *Coincidence Problem*. In this section, I want to examine more carefully what it is that's supposed to be coincidental in light of these commitments. Roughly speaking, it is some reliable degree of overlap between the intuitions or beliefs which have been favoured by selection and those which are true. I want us to try to be a little more exact.

The rough formulation I just offered is ambiguous in one important respect: it is not entirely clear which are 'the intuitions or beliefs which have been favoured by natural selection'. We can draw the following distinction. Let the *initial fund* refer to those ethical intuitions/judgments bequeathed to us by evolution, considered prior to the influence of other causal factors, including rational reflection. The initial fund represents something like a 'first draft' of the moral mind. Then, let the *derived fund* be those intuitions or beliefs which result and/or survive when the initial fund is adjusted for coherence in the pursuit of reflective equilibrium. We might suppose there are other ways in which the initial fund can be altered in light of rational reflection. For the sake of argument, I am here going along with Street's assumption that rational reflection is limited to systematizing the evaluative judgments resulting from our evolutionary inheritance.

We should then ask: Is the overlap which cannot be merely coincidental supposed to involve the reliability of the initial fund or the derived fund? I take it that the reliability of our present beliefs is what really matters to us. Since we are morally reflective creatures, these will be members of the derived fund. However, it's the initial fund that is the primary product of past selection-pressures. Furthermore, it seems we could concede that the initial fund contains a majority of falsehoods without thereby

being committed to the same view regarding the derived fund. This, I think, puts a question mark over the seriousness of the *Coincidence Problem.*

It might sound doubtful that the derived fund could be reliable if the starting fund is not. Street (2006) writes that on the latter assumption (corresponding to the first horn of her dilemma), "all our reflection over the ages has really just been a process of assessing evaluative judgements that are mostly off the mark in terms of others that are mostly off the mark. And reflection of *this* kind isn't going to get one any closer to evaluative truth" (124). What Street says here is simply false. By adjusting and revising our beliefs in the pursuit of greater coherence, it's possible to move from a set of beliefs containing a majority of falsehoods – even radical falsehoods – to a set containing largely true beliefs.

Firstly, there's the following possibility. A person might have a set containing a majority of beliefs which are false, but which approximate the truth. The set might fail to exhibit some appropriate degree of coherence. In order to render the set suitably coherent, a process of mutual adjustment may be required among its members, converting sufficiently many false beliefs to true beliefs. Consider the following toy example. I believe with some confidence that all human beings have the right not to be killed except in cases of extreme emergency. I believe somewhat less confidently that it is permissible to kill non-human animals even if nothing of great moral importance is thereby achieved. Finally, I believe with certainty that species-membership is morally irrelevant. Upon noting the incoherence of these beliefs, I might revise my belief about the moral protections afforded to humans downwards to some extent and my belief about the moral protections afforded to animals upwards to a greater extent.[6] I should then think that my initial beliefs about moral status were mostly false: I was right about the irrelevance of species-membership, but I was wrong about the

[6] Cf. McMahan (2002: 228-232).

strength of the moral protections afforded to human beings and wrong to an even greater extent about the weakness of the moral protections afforded to non-human animals. Still, I wasn't that far off: my initial beliefs were not radically false.

The pursuit of coherence can also lead from a set containing a majority of falsehoods to one with a majority of truths, even if the majority of the initial beliefs are arbitrarily far from the truth.[7] Suppose that a proper subset containing a minority of our beliefs exhibits a high degree of coherence, but the remainder do not cohere at all: neither with one another, nor with the members of the subset. Then we might be justified in revising our beliefs in such a way that we retain only the members of the subset. In this way, we could move from a set containing a majority of radical falsehoods to a set with a majority of true beliefs. As an analogy, Huemer (2008) asks us to consider a case in which a detective receives six reports about the license-plate of a getaway car used in a robbery. Suppose that two of the witnesses independently report the license-plate as 'X7841A'; the others report four wildly different license plates. In light of the agreement amongst the two reports and the disarray among the remainder, the detective would seem justified in believing that the license plate is X7841A. In that case, she should also believe that most witnesses were reporting something false - indeed something not even approximately true.

It is in principle possible, then, that we could justifiably retain our present ethical beliefs while conceding that the initial fund was thoroughly saturated with error. In light of this, we might wonder whether it need be all that implausible to insist that, coincidentally, evolution by natural selection put us in a position where our present moral beliefs are by and large accurate. We can combine this with the concession that the initial fund was shot through with distortions, just as one might expect given *Functional Truth-Irrelevance*.

[7] Ironically, this point is made by Huemer (2008).

To put point on this issue, I'd like us to consider once more Jonathan Haidt's *Moral Foundations Theory* and its application to the liberal/conservative divide. Haidt's work suggests that liberals are committed to viewing the initial fund as quite unreliable. Thus, it's not merely possible that rational reflection could move some people to a point where they should think the initial fund was suffused with error: it appears this has actually happened.[8]

According to *Moral Foundations Theory*, we recall, the moral mind is equipped with five innate modules, each of which predisposes us to moralize a distinct domain of social action. Political liberals are said to do almost the entirety of their moral thinking by reference to the first two domains, regarding the others as sources of prejudice and moral backwardness. Conservatives, by contrast, rely freely on all five foundations. Among philosophical moral theories, we may find even less of the innate foundations in play. Certain forms of *Utilitarianism* may be thought to rely on the *Welfare* foundation alone. Haidt (2012) describes Bentham as the proponent as "a one-receptor morality." (121)[9]

Haidt regards the narrowness of liberal morality as regrettable. In light of our discussion, we might think there's something attractive about it. In largely abandoning three of the five moral foundations, liberals aren't committed to viewing natural selection as an especially reliable guide to moral value: quite the opposite. Liberals, we might think, can view natural selection as having been about as bungling in its construction of our moral outlook as we might expect given *Functional Truth-*

[8] For evidence that liberal views result from greater rational reflection see Eidelman et al. (2012).

[9] Haidt's interpretation may be challenged. Kymlicka (1989) argues that there are two kinds of utilitarian, one deontological, the other teleological. Deontological utilitarians regard *Utilitarianism* as the proper interpretation of the duty to treat all persons as equals. Bentham's slogan - 'Each to count for one and none for more than one' - thus represents a requirement of fairness. Kymlicka regards not only Bentham, but also Harsanyi, Sidgwick, and Singer as deontological utilitarians. Teleological utilitarians, by contrast, are said to found their moral outlook directly on the desirability of maximizing welfare. Clearly, deontological utilitarians do not neglect considerations of fairness: their moral outlook derives from a concern for fair treatment. Thanks to Krister Bykvist for help on this point.

Irrelevance.[10] We might wonder, then, whether liberals come under any substantive pressure to revise their moral outlook in light of the *Coincidence Problem.* Perhaps it's a problem for conservatives only. If we grant Haidt's claim that *Utilitarianism* is a one-foundation morality, utilitarians may be even better placed.[11]

Someone might insist that even liberals and utilitarians are committed to a correlation between accuracy and adaptive advantage in the initial fund that's just too astonishing to be merely coincidental. That certainly doesn't strike me as obvious. Even supposing that natural selection could only favour the evolution of a true belief as a matter of coincidence, it doesn't seem unbelievable that it should do so occasionally: someone who thought natural selection had got it right in just 1 out of 1000 cases wouldn't be committed to a striking correlation. It's not intuitively clear where the degree of accuracy in the initial fund might become too great to be regarded as purely accidental. Someone might insist that the cut-off is at just the point where reliability in the derived fund becomes possible. In light of the compatibility of a reliable derived fund with a wildly off-track initial fund, we shouldn't accept this claim without further argument.

These reflections at least partially unsettle the intuition on which the *Coincidence Problem* trades. Having distinguished the derived fund from the initial fund and noted the compatibility of unreliability in the latter with reliability in the former, it becomes clear that the seriousness of the *Coincidence Problem* can't be adequately adjudicated except in light of more careful reflection. Having now considered what sort of correlation we are supposed to think would have to be coincidental in light of

[10] Cf. my discussion in section 2.2.3 of chapter 2.

[11] Besides Huemer, the thought that a utilitarian moral outlook might be invulnerable to evolutionary debunking arguments has attracted a number of authors, drawn by the thought that the tenets of *Utilitarianism* derive support from a form of rational insight that is suitably independent of our evolutionary inheritance (de Lazari-Radek & Singer 2012; Greene 2008; Singer 1981, 2005, 2006). This view has been criticised by Kahane (2011: 119-120) as unrealistic (see also Berker 2009, Tersman 2008). On the present line of thinking we need no appeal to any spooky powers of rational insight. Teleological utilitarians may remain reasonably unfazed by the *Coincidence Problem* simply because they are not obliged to regard the starting fund as minimally reliable.

Functional Truth-Irrelevance, we should proceed to examine more carefully what it means to describe some state of affairs as *coincidental*. As I'll make clear in the next section, this isn't altogether easy to pin down, with considerable room for confusion and misdirection. Once we're clearer on this matter, we'll be better positioned to say when positing a coincidence might be unacceptable, and whether it's unacceptable in the case at hand. These issues are taken up in sections 5 to 8.

4. Coincidences

In this section, we'll begin by considering more exactly what a coincidence is supposed to be. With this complete (or as complete as need be), we'll then consider whether a coincidence has to be posited by the realist who accepts the falsity of the *Tracking Account*, and whether the anti-realist is able to escape the same commitment. The answers will be 'yes' and 'no', respectively.

4.1 Conceptual analysis

In trying to understand more exactly what it means to describe something as a coincidence, I want to focus on the relationship between coincidence and probability. It's sometimes proposed that nothing is properly called a coincidence unless it is sufficiently unlikely to have occurred. Hart and Honoré (1959) write:

> We speak of a coincidence whenever the conjunction of two or more events in certain spatial or temporal relations is (1) very unlikely by ordinary standards and (2) for some reason significant or important, provided (3) that they occur without human contrivance and (4) are independent of each other. (74)

However, not everyone uses the term in this way. David Owens (1998) proposes that we define 'coincidence' purely in terms of Hart and Honoré's fourth condition: "A coincidence is an event which can be divided into components separately produced by

independent causal factors." (13) On this interpretation, coincidences do *not* have to be chancy or surprising. Owens suggests that Hart and Honoré capture the term in its ordinary meaning and that his represents a reforming definition, designed to emphasize certain theoretically interesting links between the concepts *coincidence*, *explanation*, and *cause*, such as the fact that coincidences are often treated as having no explanation. Owens' definition is nonetheless highly faithful to our linguistic intuitions. Suppose I were to say: 'I pray every afternoon for the sun to rise the next day, and it does rise; and that is not simply a coincidence.'[12] The implication here is that my prayer somehow controls the rising of the sun. Obviously, that's false, and my statement correspondingly outrageous. By contrast, it seems tangential to our acceptance or rejection of my statement whether it was unlikely or surprising that sunrise would follow my prayers. (Obviously, it was not.)

In the same vein, we might consider some of the examples noted by Marc Lange (2010) in his discussion of mathematical coincidences. According to Philip Davis (1981), it is just a coincidence that 9 is the 13th digit in the decimal expansion of both π and *e*. According to David Corfield (2004), it is just a coincidence that both 13 and 31 are prime. It is difficult to see in what sense these facts could be counted as improbable or surprising. Of course, mathematical facts are not (and do not describe) events whose causes may be independent or overlapping. In that sense, Owens' definition also falters here. However, as Lange (2010: 317) suggests, we can account for the existence of mathematical coincidences by way of a definition modelled on Owens', which speaks in terms of *facts* rather than *events* and *explanation* rather than *causation*: a conjunction of mathematical facts is counted as coincidental insofar as the conjuncts are suitably explanatorily independent of one another. This requires, of course, that we have available some suitable notion of mathematical explanation.

[12] I base this example on a case discussed by Owens (1998: 8).

Arguably, the concept *coincidence* does not admit of any neat definition by way of necessary and sufficient conditions. Virtually no concept does.[13] A more fruitful approach may be to think of the concept as constituted by a *prototype*: a list of weighted features used to determine category membership.[14] I suspect that Hart and Honoré capture the features summarised under the *coincidence*-prototype. However, Owens' reforming definition ultimately provides a better approximation of our ordinary meaning, since the weighting of these features seems to strongly favour explanatory independence as the key deciding factor. Explanatory independence is in this sense the heart of the vernacular *coincidence* concept.

Insofar as we mean to be more exact in our talk, I would suggest that we treat a coincidence as being simply a conjunction of facts whose conjuncts are explanatorily independent of one another: neither fact figures in the explanation of the other, and there is no relevant[15] explanatory factor shared by the members of the conjunction. From here, I'm going to use the term 'coincidence' under this purely explanatory definition. It's not especially important to me that we adopt this convention, however. What *is* important is that we not get confused. We should take care to keep one mental file for the question of whether there might be some suitable explanatory connection between the (mind-independent) ethical facts and the existence of certain selection pressures, and a separate file for the question of whether such an alignment might be improbable or surprising. We shouldn't presume that the two questions must be

[13] See Murphy (2002).
[14] Rosch (1973).
[15] Typically, the sorts of things which are described as coincidences will share some remote explanatory factor(s). For example, I share my birthday with Peter Singer, and this, we would say, is just a coincidence. But these events are not *completely* unrelated. For example, they have the Big Bang as a common cause. To make sense of this, we need to attend to the pragmatics of explanation. In certain contexts, factors which contribute to the explanation of a certain phenomenon are nonetheless rightly ignored. For example, if a student is asked whether there are any common causes of the First and Second World War, she would do well to cite German nationalism, but should not cite the origin of humanity in the East African Rift Valley. The conjunction of two facts which share certain explanatory factors may be rightly described as a coincidence if we are properly ignoring those explanatory factors shared by the conjuncts.

answered in the same way. I will argue that they should, in fact, be answered differently.

4.2 Is a coincidence required?

In deciding whether an alignment between accuracy and adaptive advantage would have to involve a coincidence, Street *et al.* seem to focus on the existence of suitable explanatory connections linking the right values to greater relative fitness. Hopefully, this was brought out in section 2. Thus, the explanatory conception of coincidence that I've just outlined seems to fit the terms of the debate. Using this conception, let's now ask whether it would really have to be a coincidence if natural selection has favoured the evolution of ethical intuitions and judgments that are also objectively correct. Having decided this question, we'll then consider whether rejecting *Meta-Ethical Realism* in favour of *Constructivism* could make any difference to our need to posit a coincidence of this kind.

4.2.1 Is a coincidence required, assuming Meta-Ethical Realism?

One prominent objection that has been raised against Street *et al.* is that they ignore the possibility of indirect explanatory links.[16] In other words, they consider only two ways in which to connect accuracy and adaptive advantage: either the accuracy of certain ethical intuitions/beliefs explains why they have been selected for, or else it's the other way round. There is also the possibility that one and the same set of factors might explain *both* why certain ethical norms have proven adaptively advantageous *and* why those norms are correct. Thus, it needn't be a coincidence, even for the realist, that certain norms have been both reproductively advantageous and correspondent to the moral truth: there may be an indirect explanatory connection linking the two.

[16] See Brosnan (2011), Enoch (2010a), Schafer (2010), Skarsaune (2011), Wielenberg (2010).

I don't think it's especially damning that this possibility has been neglected. The proposal succeeds only in relocating the point at which a coincidence must be posited. Suppose that one and the same factor explains why certain ethical norms have proven adaptively advantageous *and* why those norms are correct. Then we should ask: is it just a coincidence that the factors which have made norms fitness-raising are also the very same factors which make norms correct?[17] Suppose, for the sake of argument, that (*a*) the tendency of certain norms to promote a peaceful, cooperative social life explains why such norms have raised the relative fitness of groups upholding them and (*b*) the tendency of certain norms to promote a peaceful, cooperative social life makes them correct. Then, we must ask: is the conjunction of (*a*) and (*b*) just a coincidence? Presumably, if we are otherwise impressed by the *Coincidence Problem*, we won't be inclined to suppose that (*b*) explains (*a*) or vice versa. To avoid postulating a coincidence, we might then posit yet another third factor, which accounts for both (*a*) and (*b*). This would just displace the problem one step again. Unless we're happy to countenance an infinite regress of third-factor explanations, the appeal to indirect explanatory connections fails to rule out that a coincidence had to obtain. For this reason, I think Street (2011) is correct to insist that "the tracking account is the only (non-coincidence-positing) way of seeing how evolutionary forces could have pushed our values *toward* independent normative truths" (13).

However, reflecting on the line of argument just sketched might also raise suspicions that it needn't be unacceptable to posit this kind of coincidence. After all, barring infinite regresses, the explanatory buck always has to stop somewhere. Unless we suppose that all facts ultimately derive their explanation from a single brute fact,

[17]Enoch (2010a: 433) notes this problem, but his response involves a switch of mental files: he argues simply that this deeper alignment is not plausibly regarded as having a low objective probability. Probabilistic issues of this kind are addressed in section 6, where I argue that they are insufficient to defuse the *Coincidence Problem*.

we should expect to find explanatorily unrelated pairs as we delve deeper and deeper into the reasons why.

4.2.2 Is a coincidence required, assuming Meta-Ethical Constructivism?

To add to our doubts that the realist faces any especial problem, I'm now going to argue that constructivists do *not* avoid having to say that only by a coincidence has natural selection favoured the evolution of ethical intuitions/beliefs that happen also to be true. I believe that Street's argument to the contrary is flawed.

If there's to be some explanatory connection between accuracy and adaptive advantage, it seems there are only three possibilities: certain beliefs have been selected for because they are true; certain ethical beliefs are true because they were selected for; or there exists some indirect explanatory connection. The first possibility is ruled out by *Functional Truth-Irrelevance*; the third doesn't avoid the need to posit a coincidence, as we've just seen.

This leaves only the second option: certain ethical beliefs are true because holding beliefs with similar contents raised the relative fitness of our ancestors. If the constructivist is to avoid the commitment to a coincidence, she must take this view. However, the view seems quite implausible. On the face of it, the fact that a belief was selected for in the past does nothing to make it true. If belief in retributive punishment has been favoured by natural selection, this doesn't help to make retributive punishment justified. If the constructivist has to offer a theory which says otherwise, we may have trouble seeing *Meta-Ethical Constructivism* as a satisfactory response to the *Coincidence Problem*. If it sounds like I'm merely begging the question here, I should point out that I have trouble seeing *Constructivism* as even consistent with a view on which past selection explains why certain beliefs are true. According to *Meta-Ethical Constructivism*, a belief of the form *p is a reason for S to φ* is true iff and

because a belief with that content coheres suitably with *S*'s other normative judgments.[18] In that case, it would appear, the truth of a self-directed normative belief is simply due to its coherence with one's other beliefs; past selection doesn't enter into it.

You may think I am missing something obvious here. According to the constructivist, my belief in some first-personal normative proposition, *p*, is correct because I would accept *p* in reflective equilibrium. But the fact that I would accept *p* in reflective equilibrium is explained, one might think, by the fact that I belong to a species in which the disposition to accept *p* has been favoured by natural selection.[19] Therefore, the fact that a disposition to accept *p* has been favoured by natural selection explains the correctness of my normative verdict: there is no coincidence. This, I take it, encapsulates Street's reasoning on the matter.[20] The inference just sketched is faulty, however, for reasons I'll now explain.

In many cases, we aren't altogether exact about the sorts of things we want explained, nor the things we invoke to do our explaining. In such cases, the use of explicit contrast-classes can help. Consider Willie Sutton's famous reply to the question of why he robbed banks: "Because that's where the money is!" In one sense, Sutton succeeded in explaining why he robbed banks, but he didn't answer the question posed to him. We can use contrast-classes to precisify the explananda and diagnose what went wrong. Sutton managed to explain why he robbed banks *as opposed*

[18] Street (2008b).

[19] This may be doubted. Some philosophers of biology hold the view that natural selection cannot explain the properties of individuals, only populations. For this view, see Lewens (2001), Pust (2001b, 2004), Sober (1984, 1995). For objections, see Neander (1988, 1995a, 1995b), Matthen (1999, 2003), and Stegmann (2010). This view strikes me as plausible when restricted to genetic evolution, because origins essentialism entails that a person cannot exist without their lineage (Pust 2001b). However, the view strikes me as highly implausible when applied to cultural evolution, where origins essentialism creates no similar issue. Cultural selection is scarcely touched on in the debate among philosophers of biology, but arguably plays a prominent role in the evolution of morality: see chapter 1.

[20] See, *e.g.*, Street (2006: 152-154)

to (say) *libraries*, but we wanted to know why he robbed banks *as opposed to staying on the straight and narrow.*[21]

According to *Meta-Ethical Constructivism*, my belief in some first-personal normative proposition, *p*, is correct because I would accept *p* in reflective equilibrium; the fact that I would accept *p* in reflective equilibrium is explained by the fact that I belong to a species in which the disposition to accept *p* has been favoured by natural selection; therefore, it seems, the fact that a disposition to accept *p* has been favoured by natural selection explains the correctness of my normative verdict. This inference seems cogent, but the conclusion can easily strike us as mistaken. As I've noted, it doesn't seem that past selection makes some belief true as opposed to false, even assuming *Constructivism*. We can use contrast-classes to diagnose how the inference goes wrong. If we are asked to explain why my belief in some normative proposition *p* is true, the relevant contrast is my belief in *p* being false: on this construal, the explanandum is formulated so that it builds in the presumption that I believe *p*. What must be explained is a difference between my belief being one way (true) as opposed to another (false). Nothing which merely explains why I believe *p* as opposed to not believing *p* properly contributes to the explanation of my believing *p* truly: any such explanatory factor is screened off. On the constructivist's picture, my belief in *p* is true (as opposed to false) because it would be endorsed in reflective equilibrium. But in what sense do previous selection pressures favouring belief in *p* explain the fact that I would accept *p* in reflective equilibrium? They do so, I presume, only by virtue of explaining why I believe *p*, as opposed to not believing *p*. In that case, they cannot explain why my belief in *p* is true as opposed to false, since that explanandum presumes that I believe *p*.

[21] Some philosophers have claimed that explanations are always, if only implicitly, contrastive: that to explain why *p* is always to explain why *p, rather than q* (Garfinkel 1981; van Fraassen 1980). I don't wish to commit myself to anything quite so strong; I want merely to register that the use of contrast classes is a reliable means of clearing up explanatory ambiguities (Lewis 1986; Lipton 1991).

Constructivists are no better off than realists, then, when it comes to their ability to postulate suitable explanatory links between the direction of selection and the truth of the ethical intuitions/beliefs selected for. True beliefs weren't selected for, selection doesn't determine truth, and third-factor explanations are no help, we've seen. A coincidence must be involved, whether we are realists or not.

We might still feel some resistance to this conclusion. It may seem obvious that the *Coincidence Problem* must be less of an issue for the anti-realist. Guy Kahane (2011) has argued that evolutionary explanations simply cannot challenge the reliability of our ethical beliefs if we assume some form of *Meta-Ethical Constructivism*. He writes:

> anti-objectivist views claim that our ultimate evaluative concerns are the source of values; they are not themselves answerable to any independent evaluative facts. But if there is no attitude independent truth for our attitudes to track, how could it make sense to worry whether these attitudes have their distal origins in a truth-tracking process? (112)

We don't have to resist my conclusions in this section in order to agree with Kahane's point. If we're convinced that the *Coincidence Problem* must be less of a problem for the anti-realist, we should infer that there is some important aspect of the problem that we have yet to consider. There is. As I've said, we should take care to keep one mental file for the question of whether there might be some suitable explanatory connection between the ethical facts and the direction of selection, and a separate file for the question of whether their alignment might be improbable or surprising. We should consider the first file closed. This leaves open whether there might be some difference between *Realism* and *Constructivism* when we turn to the second file. That's what I now propose to do.

5. Probability and Surprise

There's an additional reason why probabilities have to enter the picture. Huemer (2008) says: "if the values toward which natural selection biased us coincided with the objectively correct values, this would be sheer coincidence. Such a coincidence cannot reasonably be expected." (377) But there is no general ban on believing in coincidences. You can happily go ahead and believe that 9 is the 13th digit in the decimal expansion of both π and *e*. Here, no one thinks 'such a coincidence cannot reasonably be expected', or even that the need to posit a coincidence like this provides any reason to doubt the theories which specify the values of π and *e*. So why can't we happily think that, as a matter of coincidence, natural selection favoured the evolution of true ethical beliefs? To answer this question, I'll argue, we need to adopt a Bayesian perspective on the *Coincidence Problem*. Once we adopt this perspective, we'll also be able to see how anti-realists enjoy some form of advantage over realists.

When it comes to the issue of why a coincidence can't reasonably be expected, we get little help from the authors discussed in section 2. Huemer is silent on the matter. Rosenberg (2011) writes simply: "A million years or more of natural selection ends up giving us all roughly the same core morality, and it's just an accident that it gave us the right one, too? Can't be. That's too much of a coincidence." (209) There's nothing here beyond an appeal to intuition. Street (2006) may be offering some approximation of an argument in the following:

> This degree of overlap between the content of evaluative truth and the content of the judgements that natural selection pushed us in the direction of making begs for an explanation. Since it is implausible to think that this overlap is a matter of sheer chance … the only conclusion left is that there is indeed some relation between evaluative truths and selective pressures. (125)

It may be the sentence beginning 'Since it is implausible…' is intended to support the preceding statement via something like the following argument: Unless there is some suitable explanatory connection between the mind-independent normative facts and

the relevant selection pressures, any alignment between the two must have been chancy, and thus should not be expected to have occurred. If this captures Street's reasoning, it implies that she's confounding the issues I've insisted we keep separate, assuming that the alternative to providing an explanation for the overlap is to suppose the overlap was improbable.

Street occasionally describes the overlap between the ethical judgments favoured by natural selection and those we think are true as *striking*. The word 'striking' does a lot of work here, I think. Not all coincidences are striking, but those that are striking invite deep suspicion. It doesn't seem striking that 9 is the 13th digit in the decimal expansions of both π and *e*. Nobody is surprised or astonished by that. By contrast, it would be astonishing if two people coincidentally ran into one another every 13th of July for three decades running. If we found that two people had run into one another repeatedly in this way, we'd strongly suspect that they were coordinating their activities in some way. The hypothesis that it was all just a big coincidence would be accepted only as a last resort.[22]

We can make progress on the *Coincidence Problem*, then, by analysing what this property of 'strikingness' amounts to. Luckily, there exists a small philosophical literature devoted to this issue.[23] Within this literature, the most widely respected account is due to Paul Horwich (1982).[24] Like the other parties to the debate, Horwich operates within a Bayesian framework, in which probabilities represent the degrees of confidence of an ideally rational agent, rather than objective chances existing in nature.[25]

---

[22] I borrow this example from Field (1998).
[23] See Harker (2012), Horwich (1982), Schlesinger (1991).
[24] Manson & Thrush (2003) provide a list of endorsements for Horwich's account; they also suggest that Horwich's view was anticipated by Ramsey and von Mises.
[25] Ultimately, the differences between Horwich, Harker, and Schlesinger do not matter to my argument in the remainder of this chapter, as I'm going to focus on the plausibility of (i) (see below) and no one disputes the first of Horwich's conditions

I'm now going to set out Horwich's approach to the nature of surprises, and then show how it can be applied to the *Coincidence Problem*. If the claims that should be made in applying Horwich's framework to the *Coincidence Problem* hold up, we'll have a basis on which to say that the coincidence is of a sort that can't reasonably be expected. In addition, the framework will allow us to recover an advantage for *Constructivism* over *Realism*, for reasons I'll explain.

The first condition on an event being surprising, according to Horwich, is that its occurrence would be assigned a very low prior probability given one's beliefs about the relevant circumstances. Thus, if I have a coin believed to be fair, I should assign a very low prior probability to obtaining heads 100 times in a row - an outcome that would be very surprising. Letting C be the believed proposition *The coin is fair* and E be *The coin is tossed 100 times, yielding heads every time*, we write this as: $\Pr_{old}(E|C) \approx 0$. This condition, however, is insufficient. We should assign equally low prior probability to any proposition describing the coin as being tossed one hundred times and yielding any particular sequence of roughly equal heads and tails (*e.g.*, THHTHTTHTTHTTTHTHHHH…). No sequence of that kind will surprise us, however. As a second condition, Horwich proposes that there should be some rival interpretation of the relevant set-up, which is not already assigned a sufficiently low prior probability, and relative to which the surprising outcome is in fact antecedently probable. Thus, let K be the proposition *The coin is heavily biased towards heads*. Supposing that K was not regarded as being too implausible at the outset, the following condition ought then to be satisfied: $\Pr_{old}(K)\Pr_{old}(E|K) >> \Pr_{old}(C)\Pr_{old}(E|C)$. It follows that $\Pr_{old}(C|E) << \Pr_{old}(C)$.[26] Therefore, on learning E, $\Pr_{new}(C) << Pr_{old}(C)$.

[26] See Horwich (1982: 101-102) for the derivation.

Here's how we can apply this model to the *Coincidence Problem*. Street thinks the overlap between the ethical judgments favoured by natural selection and those we think are true would be striking if the ethical facts were constitutively independent of our evaluative attitudes. This implies her acceptance of something like the following:

Abbreviations:

R = the reliability of our evolved ethical beliefs (derived fund)

FTI = *Functional Truth-Irrelevance*

MR = *Meta-Ethical Realism*

MC = *Meta-Ethical Constructivism*

B = our background knowledge concerning the evolution of humanity

(i) $\Pr_{old}(R \mid FTI \wedge MR \wedge B) \approx 0$;

(ii) $\Pr_{old}(R \mid FTI \wedge MC \wedge B) >> 0$;

(iii) $\Pr_{old}(MC)$ is not so very low relative to $\Pr_{old}(MR)$, such that, given (i) and given (ii):

(iv) $\Pr_{old}(R \mid FTI \wedge MC \wedge B)\Pr_{old}(FTI \wedge MC \wedge B) >>$

$\Pr_{old}(R \mid FTI \wedge MR \wedge B)\Pr_{old}(FTI \wedge MR \wedge B)$.

Are these claims true? Let's focus on (i) and (ii). Although some philosophers might assign a very lower prior to *Constructivism*, I think it would be quite unreasonable to deny (iii) and (iv) if these first two premises hold up. Furthermore, by reflecting on these premises we'll be able to gain a sense of the sort of advantage for *Constructivism* that might be on offer in light of this Bayesian re-interpretation of the *Coincidence Problem*.

Letting 'our evolved ethical beliefs' denote those beliefs which result/survive given our attempts to bring the initial fund into reflective equilibrium, (ii) seems very

plausible: *Meta-Ethical Constructivism* should lead us to expect those beliefs to be reliable, at least if they are first-personal reasons-ascriptions. After all, the constructivist treats coherence as the ultimate arbiter of truth for judgments of this kind. Furthermore, if a set of first-personal ethical beliefs exhibits sufficient coherence, *Constructivism* should lead us to assign high confidence to its members' reliability regardless of their distal origin: in this way *Constructivism* screens off aetiological considerations from *R*. This, I take it, is the point of Kahane's rhetorical question: 'if there is no attitude independent truth for our attitudes to track, how could it make sense to worry whether these attitudes have their distal origins in a truth-tracking process?'[27] Thus, $\Pr_{old}(R \mid FTI \wedge MC \wedge B)$ should be high.

Realism, of course, does not count coherence as sufficient for truth. It is conceptually possible, given *Realism*, that the ethical facts are otherwise than we take them to be, no matter how coherent our beliefs. As Russ Shafer-Landau (2003) notes: "Because realism does not see truth as constituted by even idealized attitudes taken towards some subject, realism allows for the possibility that moral truth may elude our best epistemic efforts." (225) In addition to being coherent, a reliable belief-set must lock on to a body of mind-independent normative facts. Plausibly, $(FTI \wedge B)$ offers no reason to expect that the beliefs favoured by selection will also correspond to this mind-independent domain of fact. Therefore, $\Pr_{old}(R \mid FTI \wedge MR \wedge B)$ should *not* be high, as is implied by (i). *Meta-Ethical Realism* introduces an extra condition on reliability, but the evolutionary facts provide no reason to expect our evolved beliefs to satisfy that condition.

We appear, then, to have recovered some form of advantage for *Constructivism* over *Realism* of the sort that failed to emerge when we considered only the

[27] Cf. Street: "there is no possibility of being 'off track' due to evolutionary influences … *Whatever* our initial set of [evaluative attitudes] might be, what's ultimately worth pursuing is in some way a function of those." (2011: 22 n.41)

explanatory side of things. We saw in the previous section that Street was wrong to think that constructivists are better placed to postulate suitable explanatory links connecting the truth of our present ethical beliefs to the existence of past selection-pressures favouring their evolution. The Bayesian approach I've outlined allows us to reconceive the advantage enjoyed by constructivists in probabilistic terms.

However, the advantage I've just identified is less than what's needed. I haven't given us enough to substantiate premise (i). I argued that ($FTI \wedge MR \wedge B$) provides no reason to expect reliability. Premise (i) is stronger: it says that, conditional on ($FTI \wedge MR \wedge B$), we should expect that R is false. Whether this stronger claim can be substantiated is an entirely different matter. In the remainder of the chapter, I'm going to outline the best – and, I think, only – possible justification for (i). I then argue that it fails. It's on this basis that I dismiss the *Coincidence Problem* as illusory.

In the next section, I'll assess what has already been said in the literature on the *Coincidence Problem* about the probability that natural selection would favour the evolution of reliable ethical beliefs. This, we'll see, falls short of settling the matter. It does, however, offer some valuable lessons about the irrelevance of objective probabilities. I'm going to build on these lessons in section 7, where I show how (i) can – and must - be justified.

6. An appeal to chances?

Why should your prior that a fair coin will yield heads on every one of 100 tosses be so low? Because the objective probability of that outcome - $\text{CH}(E|C)$ - is very low, and your credence should mirror this: you should defer to chance. This requirement is encapsulated in Lewis's (1980) *Principal Principle:*

The PrincipalPrinciple:

$\Pr(p \mid \text{CH}(p)=x) = x$

We might hope to justify assigning a low prior to $(R \mid FTI \wedge MR \wedge B)$ in the same way: by arguing that $\text{CH}(R \mid FTI \wedge MR \wedge B) \approx 0$. However, David Copp (2008) argues forcefully against this proposal. He insists that realists are not forced to think it was unlikely that natural selection would favour the emergence of an initial fund that allows for reliability in the derived fund.

Copp regards probabilistic issues of this kind as central to the problem posed by Street's evolutionary challenge. As Copp understands it, "The basic challenge … is to explain what it is about the moral truth such that, if the adaptive link account is correct, it is likely that our moral beliefs tend to approximate the truth." (198) He proposes to meet this challenge as follows. According to his 'society-centred' view of morality, a moral proposition, *p*, is true iff and because the moral code that best allows society to meet its 'needs' includes or implies a corresponding norm.[28] The 'needs' of a society are understood to include its need to reproduce itself, for its members to cooperative peacefully, and for peaceful and cooperative relations with neighbouring societies. Even if the *Adaptive Link Account* is true, Copp notes, the moral beliefs that we should expect to emerge from the evolutionary process would overlap considerably with those that are true according to the society-centred view: it's not a matter of chance that natural selection favoured beliefs of that kind. Furthermore, if the society-centred view is true, there's nothing chancy about that: presumably, it's a necessary truth, and necessary truths have objective probability 1. [29] Thus, it needn't be a fluke at all if the selection-pressures which have determined our ethical beliefs have favoured beliefs correspondent to the ethical truth: "if our moral beliefs are true or approximately true, this is not a matter of chance." (Copp 2008: 202)

[28] Copp (1995, 2007).
[29] Cf. Enoch (2010a), Wielenberg (2010).

Roger White (2010) has put forward a similar position. He says:

> We might deny that it is so unpredictable that evolution should produce creatures with correct moral beliefs. … [N]atural selection is likely to favor creatures with a sense of obligation toward their offspring. But necessarily, agents *do* have an obligation toward their own offspring. So it is to be expected that evolution will produce creatures with correct moral beliefs on this and a range of other matters. (589) [30]

What are we to make of this line of reasoning? Street (2008a), for one, is unimpressed. In reply to Copp, she says that his argument is "trivially question-begging": the question is begged because "our substantive normative views on how we have reason to live are … merely being taken for granted as at least roughly correct." (215) I presume she would say the very same about White.

I think Street is misreading Copp, but that her complaint is otherwise on the mark. Why is she misreading Copp? Copp's conclusion is framed as a conditional: '*if* our moral beliefs are true or approximately true, this is not a matter of chance.' A conditional statement doesn't presuppose its antecedent, so the truth and/or reliability of our ethical beliefs isn't being taken for granted here. Copp does not beg the question. The charge of question-begging looks to be in better shape when applied to White, who does seem to presume the correctness of our evolved ethical beliefs - and certainly makes no effort to hedge his conclusion. As indicated, I think Street is correct to insist that we can't simply take the reliability of our evolved ethical beliefs for granted in assessing the seriousness of the *Coincidence Problem*. However, this restriction might itself seem question-begging. Unless we've already been offered a defeater for these beliefs, why shouldn't we be able to rely on them? The reason lies in the Bayesian set-up I've described. Premise (i) concerns the *prior* probability of reliable ethical beliefs, conditional on ($FTI \wedge MR \wedge B$). Even granting that we are ordinarily

[30] Cf. Schafer (2010).

permitted to assume R, we can't do so when assessing the prior probability of R conditional on $(FTI \wedge MR \wedge B)$. That's just what it means to be talking about the *prior* probability. Thus, we can't take the reliability of our evolved ethical beliefs for granted in this context.

Let's return to Copp's conditional conclusion, which, as I've argued, isn't vulnerable to this problem. Even granting its truth, I think Copp's position poses less of a challenge to the *Coincidence Problem* than he thinks. Copp says that if our evolved ethical beliefs are in fact reliable, there was nothing especially chancy about the evolution of reliable ethical beliefs. This claim is entirely compatible with the truth of $\Pr_{old}(R|FTI \wedge MR \wedge B) \approx 0$. Your prior that some set-up will yield a given result could be very low, though you might be very confident that the set-up had a high chance of yielding that result, assuming it does. Here's an illustration. Imagine I pick a coin at random from an assortment of 1000 coins. I know that only one of these coins is heavily biased towards heads, the rest being fair. My prior that my coin will net me one hundred heads in a row should be low; but, my confidence that the coin had a high chance of netting me one hundred heads in a row, conditional on the assumption that it does, should be high.

What Copp's argument manages to do, then, is add greater pressure to the question of how we could justify assigning a low value to $\Pr_{old}(R|FTI \wedge MR \wedge B)$, without providing any positive reason to deny that the value should be low. Granting his conclusion, we can't claim that $\text{CH}(R|FTI \wedge MR \wedge B) \approx 0$ unless we assume the falsity of R.[31] Since that assumption certainly can't be be made without begging the question, this entails that (i) can't be established by straightforward application of the *Principal Principle* in combination with the premise $\text{CH}(R|FTI \wedge MR \wedge B) \approx 0$. In other

[31]Cf. Skarsaune's 'iffy' reply to Street: Skarsaune (2010).

words, the appeal to objective probabilities in support of (i) is unpromising. If we want to find a plausible basis for this premise, we have to look elsewhere.

7. The Fine-Tuning Analogy

Let's look to the philosophy of cosmology. The fundamental physical constants governing our universe appear fine-tuned to permit the existence of life. Many feel intuitively that this can't merely be a coincidence: it cries out for explanation. This is the *Fine-Tuning Problem* in cosmology. *Prima facie*, there's a close analogy with the *Coincidence Problem*. We can't just assume that the constants happen to fall into the life-permitting region, and we can't just assume that the moral beliefs favoured by natural selection happen to lock on to the mind-independent facts. This analogy is far from superficial, as I'll now show. It can be exploited to construct a plausible basis for (i) - the only possible basis, I argue. The analogy with the *Fine-Tuning Problem* leads eventually to a series of doubts about the seriousness of the *Coincidence Problem*, which I discuss in the next section.

7.1 The Fine-Tuning Problem explained

As a first step, I'll outline the *Fine-Tuning Problem* in greater detail, starting with the relevant empirical details. Our universe is governed by a number of fundamental physical parameters, such as the gravitational constant. The values of these parameters cannot be predicted from any physical theory currently known: they are in that sense arbitrary. The values actually taken by these parameters have the interesting property of appearing fine-tuned to permit the existence of life: they fall into a narrowly circumscribed life-permitting range. As Stephen Hawking (1998) notes: "there are relatively few ranges of values for the numbers that would allow the development of any form of intelligent life."

The evidence that the fundamental physical parameters are fine-tuned has been thought to support at least one of two extra-cosmic hypotheses. As Hawking says: "One can take this either as evidence of a divine purpose in Creation and the choice of the laws of science or as support for the strong anthropic principle." (130) I'm going to call the view that there is 'divine purpose in Creation and the choice of the laws of science' the *Design Hypothesis*. By the 'strong anthropic principle', Hawking has in mind the *Multiverse Hypothesis*: that there exist very many universes, varying in the values of their fundamental physical parameters, such that the probability of at least one life-permitting universe is very high.[32]

The support offered to the disjunction of the *Design Hypothesis* and the *Multiverse Hypothesis* by the fine-tuning evidence is typically formalized as a matter of Bayesian confirmation, backed by probabilistic claims akin to those I put forward in section 5.[33] The exact formalization varies from author to author. To emphasize the analogy we're interested in, we're going to write it out as follows:

Abbreviations:

K	=	our background knowledge in physics and cosmology
FT	=	that only if they fall within a narrow range are the constants life-permitting.
LP	=	that the constants governing our universe are life-permitting
LUC	=	that ours is a lone, uncreated universe, with no designer.
MV	=	*The Multiverse Hypothesis*
D	=	*The Design Hypothesis*

(a) $\Pr_{old}(LP \mid K \wedge LUC \wedge FT) \approx 0$

(b) $\Pr_{old}(LP \mid K \wedge (MV \vee D) \wedge FT) >> 0$

[32] Note that this deviates from the definition of the strong anthropic principle originally introduced by Carter (1974).
[33] See Collins (2009), Holder (2004), Leslie (1989), Swinburne (1990).

(c) $\Pr_{old}(K \wedge (MV \vee D) \wedge FT)$ is not so low relative to $\Pr_{old}(K \wedge LUC \wedge FT)$, such that, given (a) and (b):

(d) $\Pr_{old}(LP \mid K \wedge (MV \vee D) \wedge FT) \Pr_{old}(K \wedge (MV \vee D) \wedge FT) >>$
$\Pr_{old}(LP \mid K \wedge LUC \wedge FT) \Pr_{old}(K \wedge LUC \wedge FT)$

From this it follows that $\Pr_{old}(K \wedge (MV \vee D) \wedge FT \mid LP) >> \Pr_{old}(K \wedge LUC \wedge FT \mid LP)$. Thus, the discovery that our universe is fine-tuned for life should significantly decrease our confidence that ours is a lone, uncreated universe.

Let's now examine the justification typically offered in support of (a). We'll start by noting that (a) is never supported by appeal to the *Principal Principle* in combination with the posit of some lottery-style cosmogony. No one thinks (a) is true because we have prior reason to suppose the universe came into being via a process in which the values of the parameters were selected at random. Where forthcoming, arguments for (a) appeal instead to the *Principle of Indifference*.[34] The *Principle of Indifference* is a means of generating priors from evidential neutrality: cases in which we have no reason to expect one outcome rather than another (without necessarily having any positive reason to assign equal expectation to either). The principle says:

The Principle of Indifference:

Necessarily, for any *S*, *P*: If *S* has no more reason to expect any cell - p_1, p_2, … p_n - in the partition of a possibility-space, *P*, *S* should assign equal probability to any cell, p_i, in the partition: $\Pr(p_i) = 1/n$.

[34] See, *e.g.*, Collins (2009), Smolin (1997), Swinburne (1990). Strictly speaking, Collins appeals to a restricted form of the principle, designed to evade problems like the *Water Into Wine* paradox, discussed in section 8.

The application to the *Fine-Tuning Problem* seems straightforward. Our background knowledge, *K*, provides no reason to expect parameter-values falling within any particular range. Given, *FT*, the range of life-permitting values is known to be very small relative to the space of conceptually possible[35] parameter values. Therefore, by the *Principle of Indifference*, the prior probability of obtaining life-permitting values should very low, unless we take on some postulate, such as *MV* or *D*, which should lead us to expect those values to emerge.

7.2 The Coincidence Problem and the Fine-Tuning Analogy

I'll now show how something like the justification just offered for (a) can be used, *mutatis mutandis*, to support premise (i). There are three steps.

First of all, we need a large possibility-space, corresponding to the vast range of possible parameter settings. In section 4, I said that, conceptually speaking, the mind-independent ethical facts could be different from what we've evolved to believe. We might think there's a whole range of ways in which they could be different. Street (2008a) certainly holds just this view. She writes:

> According to the normative realist, there are normative truths that hold independently of all our evaluative attitudes. Moreover, as a purely conceptual matter, these independent normative truths might be anything. In other words, for all our bare normative concepts tell us, survival might be bad, our children's lives might be worthless, and the fact that someone has helped us might be a reason to hurt that person in return. (208)

[35]Why *conceptual* possibility? One might have thought that we could appeal to the *metaphysical* possibility of the extant universe having different parameter-settings. However, it is not implausible that the parameter-values of a universe are among its essential properties. See Manson (2009: 277-278).

If we're inclined to go along with this reasoning, we seem to have the possibility space we need: there is a wide range of conceptually possible worlds in which the ethical facts are varied quite substantially.

As our second step, we need to identify a narrow reliability-permitting region within this space, analogous to the narrow life-permitting region in the space of conceptually possible parameter-settings. To do this, we just need to invoke the principle that selection for or against certain moral norms is independent of their truth or falsity: those ethical beliefs (or evaluative dispositions) which have been favoured by natural selection would have been adaptively advantageous whether or not they are (or lead to beliefs that are) true. This entails that in all of those conceptually possible worlds in which the ethical facts are varied from the way we've actually evolved to think they are, the direction of selection won't vary accordingly: natural selection doesn't track the truth. Since there are so many ways in which to vary the ethical facts, the upshot appears to be a tightly-circumscribed region of the possibility space in which the beliefs favoured by selection line up with the facts.

Finally, we just need to apply the *Principle of Indifference*, appealing to the fact that ($FTI \wedge MR \wedge B$) provides no reason to expect that the moral facts are one way rather than another. Given the size of the reliability-permitting region relative to the whole, we ought to assign a very low prior to the objective ethical facts coinciding with the belief-contents favoured by selection. Therefore, $\Pr_{old}(R \mid FTI \wedge MR \wedge B) \approx 0$.

There's lots to quibble about in this set-up. I'm rather suspicious of the notion of conceptual possibility, for example. I'm just not sure which moral propositions could be true, conceptually speaking. Consider Street's claim that "as a purely conceptual matter, these independent normative truths might be anything". One might think there exist *some* conceptual constraints on what the normative facts could be. For example, one might think it conceptually impossible that the moral facts involve

contradictions. Others might attach conceptual necessity to formal principles like the transitivity of betterness or '*ought* implies *can*'. Foot (1958) once claimed that there is no moral sense of 'should' relative to which someone could believe, as a basic moral principle, that people should not run around trees left-handed or look at hedgehogs by moonlight. This might be thought to imply that it's conceptually impossible for principles like this to be true.[36]

It's difficult to decide these questions without a better understanding of what is meant by 'conceptual truth' and 'conceptual possibility'. Here is one natural suggestion: *p* is a conceptual truth iff it's analytic that *p*; *q* is conceptually possible iff *q* is logically consistent with the set of all conceptual truths. We need then only to find an acceptable account of analyticity. Depending on the conception chosen, this might place very few restrictions on the space of conceptually possible ethical truths. For example, suppose we interpret analyticity as *Frege-analyticity*: a sentence is analytic iff it can be turned into a logical truth by the substitution of synonyms for synonyms. We would then regard as conceptual truths neither the transitivity of betterness, '*ought* implies *can*', nor the view that examining hedgehogs in moonlight cannot be wrong as a matter of basic moral principle. Suppose instead that we interpret analyticity according to a conception of *epistemic analyticity* on which a sentence is analytic only if no one could understand the sentence without thereby being disposed to assent to it.[37] Williamson (2006) argues persuasively that virtually nothing is analytic in this sense, including the laws of logic. On either proposal, Street's claim that there's an unlimited range of conceptually possible ethical truths appears very plausible.

As an additional quibble about the set-up I've sketched, we might worry about the use of measure terminology. In the case of the *Fine-Tuning Problem*, modern cosmology allows us to pick out relatively precise real-valued intervals in which a

[36] Cf. Shafer-Landau (2012).
[37] Boghossian (1996, 2003) champions this conception of analyticity.

given parameter is life-permitting; the size of the life-permitting range is thus Lebesgue-measurable. It's not so clear what could be meant by the 'size' of the region of conceptually possible worlds in which natural selection and the moral facts are so aligned that some suitable percentage of the beliefs favoured by selection are true. We might feel that we have some intuitive grasp of this notion, but we may also worry that there is no notion to be grasped, intuitively of otherwise: the idea of a measure over the proposed possibility-space is nonsensical.

I will return to these issues in the next section. As we'll see, they are not merely technical problems that we can hope to solve at a later date; how these issues are resolved is crucial for whether we should take the *Coincidence Problem* seriously at all. Is there any reason, however, to believe that the kind of set-up I've sketched drives the *Coincidence Problem*? There is. In a later paper, Street (2011) writes:

> as a conceptual matter, the independent normative truth could be *anything*. For all we know as a conceptual matter, in other words, what's ultimately worth pursuing could well be hand-clasping, or writing the number 587 over and over again, or counting blades of grass. But if there are innumerable things such that it's conceptually possible they're ultimately worth pursuing, … then what are the odds that our values will have hit, as a matter of sheer coincidence, on those things which are independently really worth pursuing? That the odds seem low is an understatement. (14)

This strongly implicates acceptance of the kind of set-up I've sketched.

More importantly, there is, I think, no alternative. I say this because, on the face of it, (*FTI* ∧ *MR* ∧ *B*) *does* put us in a state of evidential neutrality with respect to *R*: we are given no reason to expect that the ethical facts are as we have evolved to think they are, but we are also given no positive reason to think that they are not.[38] Assuming that (*FTI* ∧ *MR* ∧ *B*) is in this way neutral with respect to *R*, there is no way to show that $\Pr_{old}(R \mid FTI \wedge MR \wedge B) \approx 0$ except by appeal to some means of

[38] White (2010) says: "evolutionary considerations *fail to vindicate* our moral judgments, they don't provide any *further reason to doubt* the reliability of these judgments." (590)

generating priors from evidential neutrality. This, then, seems to requires reliance on the *Principle of Indifference* in combination with a roomy possibility-space in which the *R*-region is dwarfed by its complement.

The analogy between the *Coincidence Problem* and the *Fine-Tuning Problem* is thus far from superficial: they are fundamentally alike. This, I'm now going to argue, is ultimately more of a burden for the *Coincidence Problem* than a blessing.

8. … and the Lord taketh away

In tracing connections between the *Fine-Tuning Problem* and the *Coincidence Problem*, we've so far seen how the latter could be helped along by an analogy with the former. We'll now see how their kinship threatens to undermine the *Coincidence Problem*.

8.1 Problems with the Principle of Indifference

As a first point, we should note that the *Principle of Indifference* is widely rejected.[39] While counterexamples to the principle are routine in introductory courses on probability, I've found surprisingly many moral philosophers unaware that the principle is routinely assigned to the scrapheap of intellectual history. The principle is so widely rejected because there are cases in which it requires us to assign incompatible probabilities to the same outcome under equivalent descriptions.

Richard Von Mises's (1957) *Water into Wine* example illustrates the problem. We are asked to assign a probability to x, the ratio of water to wine in a glass, knowing only that it lies in the interval 1/2 to 2. At the same time, we are asked to assign a probability to x', the ratio of wine to water, knowing only that it lies within the same interval. The ranges 1/2 to 1, 1 to 3/2, and 3/2 to 2 are of equivalent size.

[39] See, *e.g.*, van Fraassen (1989): "It is true that the historical controversy extended into our century, but I regard it as clearly settled now that probability is not uniquely assignable on the basis of a Principle of Indifference, or any other logical grounds." (292)

By the *Principle of Indifference*, each should be assigned an equal probability of $1/3$ when considering the value of x. The same reasoning applies to obtaining a value for x' in the ranges $1/2$ to 1, 1 to $3/2$, and $3/2$ to 2. By the *Principle of Indifference*, each should be assigned probability $1/3$. Since x' trivially equals $1/x$, this leads to probabilistic incoherence. We assign $1/3$ probability to obtaining a value for x' in the range $1/2$ to 1; this outcome is equivalent to obtaining a value for x in the range 1 to 2, but the *Principle of Indifference* also requires us to assign a probability of $2/3$ to that eventuality.

I'd be happy to dismiss both the *Coincidence Problem* and the *Fine-Tuning Problem* on this basis. However, I don't want to put too much weight on problems of this sort, as philosophers continue to propose new solutions to the *Water into Wine* paradox and other cases like it.[40] For the sake of argument, I'm happy to go along with the assumption that the *Principle of Indifference* is problem-free. This only gets us out of the frying pan. There are other probabilistic issues afflicting the *Fine-Tuning Problem*. We'll see that these appear to apply, *mutatis mutandis*, to the *Coincidence Problem*.

8.2 Problems with infinite ranges

Much recent work on the *Fine-Tuning Problem* has focused on *the Normalization Problem* and the closely associated *Coarse-Tuning Problem*.[41] The problems arise because it appears that many or all of the parameters governing the universe are, as a matter of conceptual possibility, unbounded (in at least one direction). For example, as

[40] *E.g.*, Mikkelson (2004), White (2009).
[41] Colyvan et al. (2005); Davies (1992); McGrew et al. (2001); Manson (2000).

a matter of conceptual possibility, the gravitational constant could be any real number (greater than zero). The space of conceptually possible parameter-settings is infinite, whereas the life-permitting region represents a finite proportion of the total.

Here is why this creates problems. The *Principle of Indifference* requires us to assign a uniform probability distribution over equal regions of the total possibility space. The total space of conceptually possible parameter-values consists of infinitely many sub-regions of equal size to the life-permitting region. Given the standard axiomatisation of the probability calculus due to Kolmogorov (1950), it is impossible to assign a uniform probability distribution over a countably infinite partition of the total possibility-space consistent with the assumption that probability is a *normalized measure* in which the measure of the total is 1. This is the *Normalization Problem*. It relies on the axiom of *Countable Additivity*:

Countable Additivity:

If p_1, p_2, p_3… is a countable sequence of mutually exclusive outcomes and $\bigvee_i p_i$ is true iff at least some member of the sequence, p_i, is true, $\Pr(\bigvee_i p_i) = \sum \Pr(p_i)$.

Given the axiom of *Normality*, the probability of the disjunction of all cells in the partition must be 1. If the partition involves infinitely many cells, there's no real-valued probability such that you could assign the same probability to each cell and these sum to 1. Attempting to attach any prior probability to obtaining parameter-settings in the life-permitting region is therefore going to require a non-uniform probability distribution, in violation of the *Principle of Indifference*.

Though almost always used in mathematical treatments of probability, *Countable Additivity* is one of the more philosophically controversial of the Kolmogorov axioms. It has been thought undesirable because it implies, for reasons

just outlined, that there can't be a fair lottery over an infinite set of possible outcomes.[42] It has been suggested, therefore, that we adopt only the weaker axiom of *Additivity*, which treats the probability of the disjunction of any pair of mutually exclusive outcomes as equal to the sum of their probabilities:

Additivity:

If p and q are mutually exclusive, $\Pr(p \vee q) = \Pr(p) + \Pr(q)$.

This allows for the possibility of fair infinite lotteries. The rejection of *Countable Additivity* for *Additivity* has been endorsed by defenders of the *Fine-Tuning Problem*.[43] It can be shown, however, that credences which violate *Countable Additivity* leave one vulnerable to a Dutch Book,[44] typically regarded as the paradigm symptom of irrational credence.[45]

Even if we ignore this issue, we're still in the fire. If probabilities are real-valued, there is only one possible uniform probability distribution over an infinite partition that can be made, consistent with *Additivity* and *Normality*: zero. Any positive value would eventually sum to more than 1. Thus, relative to the unbounded space of conceptually possible parameter-settings, the probability of *any* finite region of the space is always the same no matter how great the region: zero. This creates the *Coarse-Tuning Problem*: the fact that the parameters must be fine-tuned to permit life is no better reason to assign a low prior to the lone, uncreated universe being life-permitting than the fact that there is *some* finite life-permitting region of arbitrary size. The probabilities are the same in either case. It follows that we should be no more surprised that the actual parameter-settings fall within a narrow life-permitting range

[42] See de Finetti (1974).
[43] *E.g.*, Collins (2009).
[44] Williamson (1999).
[45] See Howson (2008) for critical discussion of the Dutch Book argument for *Countable Additivity*.

than that they fall within *any* finite life-permitting range. The existence of *some* upper (and lower) bound on the space of life-permitting universes offers equally good reason to believe in the existence of a God or a multiverse. Many philosophers regard these implications as absurd.

Let me now spell out how the *Coincidence Problem* appears vulnerable to similar objections. The problem arises from the apparent lack of any bound on the possibility space used in arguing for (i). Street claims (2008a), as we've seen, that "as a purely conceptual matter, these independent normative truths might be anything." (208). She says: "as a conceptual matter … what's ultimately worth pursuing could well be hand-clasping, or writing the number 587 over and over again, or counting blades of grass. … [T]here are innumerable things such that it's conceptually possible they're ultimately worth pursuing" (2011: 14) We've seen some plausible interpretations of the notion of conceptual possibility that support this view. In that case, the difficulties we've identified for the *Fine-Tuning Problem* would appear to apply with equal force. In order to attribute any probability to the finite reliability-permitting region, we'd have to reject *Countable Additivity*. We'd have to attribute zero probability to obtaining the reliability-permitting region, and we'd then be faced with our own version of the *Coarse Tuning Problem*. In section 3, I said that if someone regarded natural selection as being coincidentally accurate in selecting moral beliefs/intuitions at a rate of 0.1%, we wouldn't think this a commitment to a striking, implausible correlation. By contrast, we might think a reliability-rate of 99.9% too good to be true. However, both the region in which 0.1% of beliefs which are selectively advantageous are true and the region in which 99.9% are true presumably represent bounded regions within the total, unbounded possibility-space, and so end up with the same probability: zero. It would be no less implausible, then, to suppose that natural selection had coincidentally been

accurate in 0.1% of cases than in 99.9% of cases. This implication appears equally absurd.

Admittedly, it is difficult to say with especial confidence that these problems apply to *Coincidence Problem* because, as I've noted, it's unclear what is meant by conceptual possibility and what could be meant by speaking of the size of the relevant possibility space. However, this should provide cold comfort for those who would convince us that the *Coincidence Problem* must be taken seriously. After all, it is no more clear that the *Coincidence Problem* avoids the issues I've noted. Absent some attempt to precisify the notions of conceptual possibility and measure in such a way that the problems associated with infinite ranges are shown to evaporate, we should be unconvinced that (i) is true.

Someone might think these problems are less than fatal even if they apply. The *Normalization Problem* and *Coarse-Tuning Problem* have hardly created a consensus in the philosophy of cosmology that fine-tuning is unremarkable. Some are happy to bite the bullets. Jeffrey Koperski (2005) suggests that the *Coarse-Tuning Problem* is just another of the counterintuitive results that arise when we regiment our thinking about infinity. Robin Collins (2009) has argued that the existence of an infinite range cannot plausibly undermine the *Fine-Tuning Problem*: this would imply that although the problem is serious given a finite range and more serious the greater the size of that range, the problem disappears altogether when the range becomes infinite. It is more plausible, Collins suggests, to think it should be greater still. Someone could well take a similar stance when it comes to the *Coincidence Problem*.

8.3 Problems with metaphysical non-naturalism

For those so tempted, here is one final respect in which they find a false friend in the *Fine-Tuning Problem*. It's plausible, I think, that if evidence of fine-tuning should lead

us to revise our cosmological beliefs, then the *Coincidence Problem* shouldn't worry us at all. On the other hand, if the *Fine-Tuning Problem* is really illusory, it's highly plausible the same holds for the *Coincidence Problem*. The *Coincidence Problem* would be without merit in either case. Let me now explain why.

The problem lies in the status of the *Fine-Tuning Problem* as a putative source of confirmation for theistic religious belief. If theism is probable given our background knowledge, *B*, then presumably *R* is to be expected, regardless of *MR* and *FTI*. Then (i) would be false. It might be surprising that God chose to impart the correct ethical beliefs into our species without selection for true moral beleifs, but that's about it. (He works in mysterious ways.) The game changes entirely as we move away from metaphysical naturalism and toward the hypothesis of divine creation.[46] To the extent that the *Fine-Tuning Problem* provides confirmation for theism, it undermines the *Coincidence Problem*.

Street *et al.* had better, then, have some means of blocking the *Fine-Tuning Problem* as a basis for confidence in theism. However, they're constrained in this respect: the *Coincidence Problem* and the *Fine-Tuning Problem* are so fundamentally similar that any issue with the latter is likely to be an issue with the former. I've pointed to a number of well-known difficulties which face the *Fine-Tuning Problem* and shown that similar problems appear to apply to the *Coincidence Problem*. Either we take problems of this kind seriously, or we don't. If we don't, we're going to need to say something else in response to the *Fine-Tuning Problem*.

This might not seem an especially onerous demand. It may seem obvious that we can just rely on the *Multiverse Hypothesis* to block any inference from fine-tuning to theism. Assuming (a)-(d), *LP* provides strong evidence for ($MV \vee D$), but this leaves open how we ought to distribute our added confidence across the disjuncts, and

[46] Cf. Plantinga (1998). See also Nagel (2012).

whether we ought to end up more confident of *MV* or *D*. Many have felt that we should strongly favour the multiverse over theism.

With respect to the degree of support offered by *LP* to *MV*, I believe that *LP* provides no support at all. White (2000) argues in defence of this claim. Assuming that the parameter-values obtaining in any member of the multiverse are independent of those obtaining in any other, the probability of obtaining values in the life-permitting range in this universe are no higher conditional on the existence of any greater number of universes. Of course, *MV* raises the probability of life-permitting values in *some* universe. However, rational updating requires conditionalizing on the strongest proposition learned, and the existence of *some* universe with life-permitting values is weaker than the actual fine-tuning evidence: that the parameters in *this* universe fall within the narrow life-permitting range.

If this reasoning sounds unconvincing in the abstract, an example will help:

The Firing Squad:

You are put to a firing squad. Twenty marksmen line up. Each is a fantastic shot: if they mean to hit you, there's little chance even one will miss. The guns go off. Astonishingly, you are still alive.

You might suspect that the marksmen all just happened to miss. You should certainly give substantial credence to the view that your survival was no accident: the marksmen had decided not to aim for you, or loaded their rifles with blanks. You would *not*, however, raise your confidence that this firing squad is but one of a collection of similar firing squads, large enough so that at least one squad is expected to miss their target. Although it raises the probability that someone would survive a firing-squad like the one you've faced, the multi-squad hypothesis receives no confirmation from the

fact of your survival, whereas the hypothesis of 'design' is rendered significantly more probable. By analogy, *LP* confirms *D*, but not *MV*.[47]

There are, of course, alternative routes by which to contest the use of fine-tuning as an argument for theism. We might insist that even if fine-tuning provides evidence for a designer, it offers little support for the hypothesis of a *theistic* designer. Hume (1779/1993) offers this reply to the old-fashioned *Teleological Argument*. Granting that the universe was designed, for all we know, Hume says, the designer may be "some infant Deity, who afterwards abandoned it" or the universe may be "the production of old age and dotage in some superannuated Deity; and ever since his death has run on at adventures" (71). Whether this line of objection is cogent hinges on the prior probability of theism *vis-à-vis* these alternative design-hypotheses. Our prior for theism conditional on the existence of a cosmic designer might be significantly higher, in which case the objection would be void.[48] Another objection that might be lodged against inferring theism from fine-tuning also turns on priors. One might suggest that the absolute prior probability of theism is very low, so its posterior probability remains very low in spite of any evidence offered by fine-tuning.

Questions of this kind about priors are notoriously difficult, so let me just note the following. According to the PhilPapers Surveys, nearly 15% of professional philosophers accept or lean towards theism.[49] I presume that most would do so regardless of the *Fine-Tuning Problem*. Let's suppose, on a very conservative estimate, that just 10% of professional philosophers would do so. Nobody, I presume, leans towards any of Hume's alternative design-hypotheses; very few accept some form of Deism.[50] Purely in light of these facts, one might think my prior for theism shouldn't

[47] See Bostrom (2002) and Manson & Thrush (2003) for objections to White.
[48] In Hume's defence, he offers this objection against the attempt to rest theistic religious belief entirely on evidence of design in nature.
[49] Bourget & Chalmers (2010).
[50] When asked to choose between theism and atheism, less than 1% of respondents to the PhilPapers survey chose 'Accept another alternative'.

be all that low, and should be far higher than for any alternative hypothesis consistent with the posit of a cosmic designer.

Assuming these admittedly cursory remarks hold up, we should then conclude that the *Coincidence Problem* is serious only if the *Fine-Tuning Problem* is not. As I've noted, this puts Street *et al.* in an uncomfortable position: the *Coincidence Problem* and the *Fine-Tuning Problem* are fundamentally alike, and any issue with the latter is likely to be an issue with the former. If the *Fine-Tuning Problem* is really illusory, it's highly plausible the same holds for the *Coincidence Problem.* At the very least, should it turn out that we can accept fine-tuning as merely coincidental, we should be much less confident of our ability to identify by intuition those things which genuinely cry out for explanation. We should then be much less confident in assuming that a coincidence cannot reasonably be expected when it comes to evolution and ethics.

9. Conclusion

My aim in this chapter has been to gradually chip away at the *Coincidence Problem*, getting us to a point where we can comfortably dismiss it as illusory, in spite of our intuitions. I began in section 3 by distinguishing the initial fund and the derived fund, noting that the latter might be reliable even if the former is not. I pointed out that contemporary liberals appear committed to viewing the initial fund as suffused with error in any case. It is unclear, therefore, whether liberals come under any pressure to revise their moral beliefs in light of the *Coincidence Problem.* Utilitarians may be even better placed. Hoping to a gain a deeper understanding of the issue, I offered an analysis of the *coincidence* concept. I argued that the concept might profitably be viewed as a prototype in which explanatory independence is very heavily weighted, but low probability also counts for something. For the sake of clarity, I suggested we adopt a purely explanatory definition. I used this definition to argue that realists and anti-

realists are equally committed to the view that only by a coincidence could natural selection have favoured the evolution of ethical beliefs which happen also to be true.

I then sought to understand the probabilistic aspects of the *Coincidence Problem*. I noted that if the overlap between accuracy and adaptive advantage is to be counted as striking or unbelievable, we would do well to interpret the *Coincidence Problem* in light of Horwich's Bayesian account of surprises. I outlined how this could be done, and how it allowed anti-realists to regain some form of advantage over realists in facing the *Coincidence Problem*. I then considered what could be said in support of the crucial claim that, conditional on *Meta-Ethical Realism*, we should expect the derived fund to be unreliable. I argued that this claim couldn't be supported by an appeal to objective probabilities. Instead, it would have to be supported by appeal to the *Principle of Indifference*, in just the way that those working on the *Fine-Tuning Problem* have sought to justify assigning a low prior to the lone, uncreated universe being life-permitting. I then pointed out that, if understood on this basis, the *Coincidence Problem* appears vulnerable to certain of the probabilistic issues which have been raised against the *Fine-Tuning Problem*: problems associated with the *Principle of Indifference* and with infinite ranges. I also argued that if evidence of fine-tuning should lead us to revise our cosmological beliefs, the *Coincidence Problem* shouldn't worry us, since we ought to assign significant confidence to theism. On the other hand, if the *Fine-Tuning Problem* is really illusory, it's highly plausible the same holds for the *Coincidence Problem*.

Under the weight of all these problems, I believe the *Coincidence Problem* can be dismissed. I believe it is unobjectionable to suppose the moral beliefs favoured by natural selection happen to align with the mind-independent facts. In general, our intuitions about the things that 'couldn't be merely coincidental' are not very reliable.

They are in error here. The *Coincidence Problem* is illusory: the titular coincidence is unproblematic.

5.

"The dissentient worlds of other people": phylogeny, contingency anxiety, and the epistemology of disagreement

1. Introduction

Over the previous chapters, I criticised a number of arguments that seek to derive debunking implications from *Functional Truth-Irrelevance* without appeal to *Phyletic Contingency*. These are the most prominent arguments in the literature on evolutionary debunking, and I hope to have shown that they all fail. I do not have a knock-down argument to show that it is impossible to construct a cogent debunking argument narrowly focused on *Functional Truth-Irrelevance*. However, the failures I've noted in the most popular and widely-discussed arguments lend support to the view that a cogent debunking argument must be found elsewhere.

In this chapter and the next, I will shift my focus to considering whether a cogent debunking argument can be constructed by appeal to the claim that our moral beliefs reflect arbitrary contingencies of our phylogeny. I argue that it can. In this chapter, I will argue for the claim that if evidence of phyletic contingency is debunking, this must be decided by reference to the epistemic significance of moral disagreement; in the next chapter, I argue that the epistemology of moral disagreement decides in favour of the view that evidence of phyletic contingency is defeating.

A key failing in much of the literature on evolutionary debunking arguments, I have noted, is that proponents of debunking arguments often fail to grasp the deeper epistemological issues at play. The previous chapters were devoted in large part to correcting this imbalance, exposing the epistemic foundations on which popular debunking arguments rely. My hope in this chapter and the next is to avoid any similar mistakes by focusing squarely on the broader epistemological issues relevant to constructing a debunking argument rooted in *Phyletic Contingency*. Thus, my argument in this chapter relies on the idea that evidence of the phyletic contingency of our moral beliefs is an instance of a broader epistemic phenomenon, which I call

contingency anxiety. This refers to the sense of unease we often feel when we discover that we hold certain beliefs due to arbitrary features of our background, such as the identity of our parents, the culture in which we were raised, our gender, *etc.* Given variation in these background features, we (or someone otherwise similar) would have believed the contrary of what we now believe. Because the background features are arbitrary, this sense of contingency makes us uneasy: we feel some inclination to lower our confidence or revise our beliefs altogether. I'll give some concrete examples of this in section 2.

This chapter is really about contingency anxiety in general, and not about phyletic contingency in particular. The epistemology of contingency anxiety is underexplored,[1] but encompasses a number of important issues in epistemology, including some hot-button results in experimental philosophy. I argue that if cases of contingency anxiety involve defeaters, this is because of the epistemic significance of disagreement; I call this the *Disagreement Hypothesis*. I simply apply this principle to the case of phyletic contingency.

Here is the plan for this chapter. In section 2, I clarify some terminological issues and give a number of cases that I think are exemplary of contingency anxiety. My hope is to impress on you the wisdom of thinking about phyletic contingency as a species of this genus. I then consider why cases of contingency anxiety might involve defeaters. In section 3, I show that certain natural answers to this question don't stand up to scrutiny, and I introduce the claim, inspired by Roger White's (2010) recent work on this issue, that whatever defeaters arise in cases of contingency anxiety are due to the epistemic significance of disagreement. I'll outline how my view differs from White's. I then note two significant hurdles to our accepting the *Disagreement Hypothesis*. The first is that, while many cases of contingency anxiety involve

[1] *Pace* Cohen (2000), Elga (ms), Kramer (2009), Schechter (ms), Sher (2001), White (2010).

disagreement, some don't, and those that do can seem equally problematic if re-described in such a way that no disagreement is present. The second problem is that if we assume that whatever defeaters arise in cases of contingency anxiety are due to the epistemic significance of disagreement, this appears to leave no work for our awareness of the influence of arbitrary background factors in terms of giving us reasons to revise our beliefs. It would be highly surprising, however, if these aetiological considerations turned out to be epistemically irrelevant: they seem so salient to our unease. In sections 4 and 5, I show that each of these problems can be successfully accommodated by the *Disagreement Hypothesis*. This is a case where the Nietzschean dictum about the things that don't kill us holds true: these phenomena, which initially seem fatal for the *Disagreement Hypothesis*, are ultimately explained by it with remarkable success. In light of its demonstrated usefulness as a framework for making sense of contingency anxiety, I conclude that we should adopt the *Disagreement Hypothesis*. In section 6 I consider its implications for the issue of phyletic contingency and evolutionary debunking arguments in ethics. The (slightly surprising) view at which we end up is that our attitude to the question of whether evolutionary explanations can serve as defeaters for our moral beliefs should be a function of our view about the epistemic significance of moral disagreement.

2. Contingency Anxiety

2.1 Definition(s)

To begin, we should be clear on some basic definitional issues. Let us say that a condition *C* is *arbitrary* relative to *S*'s belief that *p* iff *S*'s knowing that *C* obtains provides (or would provide) no *prima facie* reason for *S* to believe *p*. Thus, whether I'm wearing red socks is arbitrary with respect to my belief that Istanbul is in Turkey. Let

us then say that a condition *C* is a *background factor* relative to *S*'s believing *p* iff *C* contributes to the explanation of *S*'s believing *p*, but *C* forms no part of the *grounds* on which *S* bases her belief that *p*. Thus, my having been born in Denmark is a background factor relative to my belief that there is a Jutlandish dialect that allows you to construct a meaningful sentence with no consonants.[2]

Contingency anxiety arises, as I've said, in cases where we discover that our beliefs about some topic are contingent on some arbitrary background factor, such that varying the background would yield some contrary belief at the other end. I should make clear that, so understood, contingency anxiety does not at all exhaust the wider phenomenon whereby we come to feel that we ought to revise our beliefs in light of new information about their background. Our topic is much narrower than that. To take an extreme example, suppose I discover that I have always been just a brain in a vat. I ought now to be much less confident that there is a university in Cambridge. This, however, is not because of any worry to do with what I would have believed had I not been envatted. For all I know, I would have believed the same. This is a case in which information about background factors should lead me to reduce my confidence, but it is not a case of contingency anxiety.

2.2 Some examples

Here are some examples of the genuine article. Writing in *On Liberty* of the high confidence placed by the typical man in his own opinions and in those shared by individuals around him, Mill (1859/1991) says:

> He devolves upon his own world the responsibility of being in the right against the dissentient worlds of other people and it never troubles him that mere accident has decided which of these numerous worlds is the object of his

[2] As follows: *'e' e' u'e 'å æ ø i æ å*. Transl: 'I am out on the island in the brook.'

> reliance, and that the same causes which make him a Churchman in London, would have made him a Buddhist or a Confucian in Pekin. (23)

Mill believes, of course, that this should be troubling.

A similar position is taken by Jerry Cohen (2000). Cohen enjoyed a strongly political upbringing among working-class Marxists in Montreal. He claims, plausibly, that he is a Marxist *because* of his upbringing: had he been raised in the upper-middle-class part of Montreal, his present political beliefs would not be nearly so left-wing. And this troubles him: he feels much less confident of his political beliefs, knowing that he would not have them had he been raised differently.

Cohen's contingency anxiety extends to his beliefs about the philosophy of language. When Cohen left Canada in the 1950s, he had to choose between attending graduate school at Oxford or Harvard. Finding the prospect of leaving for Europe more romantic, he chose Oxford. He came out of the *B.Phil.* believing, like his fellow Oxonians, that certain truths were analytic, others synthetic. His Crimson counterparts came out denying the existence of any such distinction. Cohen says: "I think I can say that I believe in the analytic/synthetic distinction because I studied at Oxford. And that is disturbing. For the fact that I studied at Oxford is no reason for thinking that the distinction is sound." (18)

Continuing in the vein of such rarefied philosophical beliefs, Jonathan Weinberg, Shaun Nichols, and Stephen Stich (2001) claim that analytic philosophers' intuitions about knowledge are culturally parochial, reflecting arbitrary facets of Western social organization and/or social identity. Their work takes inspiration from Richard Nisbett's (2003) claim that Western and East Asian subjects rely on differing cognitive styles, traceable to differences in social structure and social identity, in turn reflecting long-run historical differences in the economic means of subsistence. At the level of cognition, Western subjects are said to rely to a greater extent on an analytic

mode of thought; East Asians are said to show greater reliance on a holistic cognitive style. These differing cognitive styles are taken to reflect the greater social individualism of Western culture *vis-à-vis* the greater collectivism of East Asian culture. Weinberg *et al.* take Gettier cases to involve beliefs which strongly resemble ordinary cases of knowledge but for some element of deviant causation. Since the analytic cognitive style focuses more on causation and agency while the holistic style focuses more on case-similarity, it was predicted that Western and East Asian subjects would differ in their epistemic intuitions about a standard Gettier case. This prediction was apparently confirmed.[3, 4] Because they seem contingent on arbitrary cultural factors, the authors suggest that epistemic intuitions are not apt to play the role they have come to play in analytic epistemology.

More generally, the experimental philosophy literature is littered with cases in which people's intuitive judgments appear to vary according to some arbitrary background factor: cultural differences are correlated with differing intuitions about reference;[5] people's moral and epistemic intuitions about certain cases depend on the order in which they are given;[6] some philosophical intuitions are correlated with heritable personality traits;[7] and some studies have found significant correlations between gender and philosophical intuitions.[8]

I'm sure you can think of other examples that could go on this list. I also hope that you feel, intuitively, that phyletic contingency belongs with the cases I've just discussed. The worrying thought that our moral outlook reflects parochial features of

[3] Jennifer Nagel (personal communication) points out that the Gettier case used by Weinberg et al. relies heavily on information about American car brands. As a means of determining whether epistemic intuitions are culturally variable, one might expect that selecting a vignette dense with culturally specific information of this kind would compromise the internal validity of the study.
[4] To date, there have been at least three failed replications of the results obtained by Weinberg et al. There are additional reasons to be suspicious, including the use of Gettier cases in Hindu-Tibetan epistemology dating to the 8th century. See Boyd and Nagel (forthcoming).
[5] Machery et al. (2004).
[6] Liao et al. (2012); Schwitzgebel & Cushman (2012); Swain et al. (2008).
[7] Feltz & Cokely (2012).
[8] Buckwalter & Stich (forthcoming).

our hominine phylogeny is, I take it, yet another case in which the discovery that our beliefs reflect some arbitrary feature of our background instils in us the worry that we shouldn't place too much confidence in them. The question for us, then, is what we are to make of these sorts of cases. The next section will begin to address this issue.

3. The *Disagreement Hypothesis*

My first task in this section will be to make the phenomenon of contingency anxiety seem more perplexing than it might first have appeared. To do so, I'll show that certain natural explanations for why cases of contingency anxiety might involve defeaters turn out to be mistaken.

3.1 What contingency anxiety could not be

Consider, firstly, the following principle:

Arbitrary Doxastic Defeat:

Necessarily, for any *S*, *p*: If *S* believes *p* with *prima facie* justification but knows that she (or someone otherwise similar) would have believed some contrary of *p* had some arbitrary background factor been otherwise, *S* thereby has a defeater for her belief that *p*.

This principle might seem plausible in light of some of the cases we've considered. We might also feel some attraction to the following, closely related principle:

Arbitrary Intuitional Defeat:

Necessarily, for any *S*, *p*: If *S* has the intuition that *p* but knows that she (or someone otherwise similar) would have had an intuition whose content is some contrary of *p* had some arbitrary background factor been otherwise, *S* thereby has a defeater for believing *p* on the basis of her

intuition.

Arbitrary Intuitional Defeat is very similar to a principle identified by Joachim Horvath (2010) as underlying the experimentalists' critique of philosophical intuitions:

Horvath's Principle:

"If intuitions about hypothetical cases vary with irrelevant factors, then they are not epistemically trustworthy." (448)

Adam Feltz and Edward Cokely (2012) have endorsed this principle, and it seems a good fit for remarks made by Joshua Alexander and Jonathan Weinberg (2007). However, *Horvath's Principle*, like *Arbitrary Intuitional Defeat* and *Arbitrary Doxastic Defeat*, is wrong. To see the falsity of *Arbitrary Doxastic Defeat*, consider:

Corner Shop:

Albert believes that the corner shop closes at midnight. A lover of spaghetti, Albert eats too much pasta one night, has trouble falling asleep, and goes for a walk around the block: he discovers that the corner shop is still open at 1 am. Albert realizes that if he didn't love spaghetti so much, he would still believe that the corner shop closes at midnight.

Albert's greed for pasta is an arbitrary background factor, but his knowledge that he would have believed the opposite of what he currently believes had he not loved spaghetti shouldn't shake his confidence that the shop is open past midnight. So *Arbitrary Doxastic Defeat* must be false.

Here's a case that seems to rule out both *Arbitrary Intuitional Defeat* and *Horvath's Principle*.

My Bayesian Love:

I once studied probability theory with great passion because I wanted to impress a woman who happened to be a Bayesian. Consequently, my probabilistic intuitions are fine-tuned and avoid all the common biases. I'm asked about a case involving the probability of obtaining some result from some hypothetical chance setup. My intuition is that the probability would be 0.8. I know that if I hadn't been in love with that woman, my probabilistic intuitions would have been different and I would have given some other answer.

Affairs of the heart are irrelevant to the facts of probability. Still, my awareness that my intuition would have been otherwise but for my romantic inclinations – and that intuitions about hypothetical cases can vary according to whom one finds desirable – shouldn't worry me at all.

How do these cases differ from those described in section 2? Clearly, not in respect of being cases in which some arbitrary background factor makes the difference between one's belief/intuition being one way as opposed to another. But wasn't that precisely what was supposed to be disturbing about those cases? What is really going on, then?

Here is one possible reply. There's some inclination to say that what is worrying about the case discussed in section 2 is that one could be right only as a matter of luck. Mill hones in on the fact that "mere accident has decided" what our beliefs should be, and we might well think of "mere accident" as equivalent to luck. It might be thought that what differentiates *Corner Shop* or *My Bayesian Love* from the cases in section 2 is that these involve a different and less disturbing form of epistemic luck. There are many varieties of epistemic luck, after all. Although knowledge excludes some forms of luck, it does not exclude all.[9] Similarly, we might think that awareness that one could be right about some issue only as a matter of luck might be

[9] See Pritchard (2005), Unger (1968), Zagzebski (1999).

defeating, but only for certain forms of luck – forms present in the cases discussed in section 2, but not in those discussed above. This is what we should expect if our instincts are right in identifying epistemic luck as significant to the cases from section 2.

There are two problems for this proposal, however. The first is that it is not clear that there is any form of luck that would allow us to differentiate between the two kinds of case. Consider two varieties of luck that are widely agreed to be compatible with knowledge: *evidential luck* and *capacity luck*. The former occurs when a person is lucky to have the evidence that she relies on, the latter when she is lucky to have the relevant cognitive abilities. *Corner Shop* may be thought to involve evidential luck, and *My Bayesian Love* might seem to involve capacity luck. It is not clear, however, that there is any different element of luck involved in the cases discussed in section 2. To the extent that these involve luck, the luck seems to relate to having been granted the right evidence or the ability to respond to it correctly.

The second problem is that luck seems inessential to the cases discussed in section 2. It cannot easily be attributed in some. Plausibly, a state of affairs is lucky only if it obtains in the actual world but not in most nearby worlds: anything lucky could easily have failed to occur.[10] Because personality and gender are to a great extent down to genetic origin, which is plausibly thought essential,[11] it is doubtful that one's personality or gender is a matter of luck. However, gender and heritable personality traits are potential sources of contingency anxiety, as we saw. These points apply with even greater force to the case of phyletic contingency: there is presumably no close possible world in which I belong instead to a distantly-related species.

Once we see that luck is inessential to contingency anxiety, we can envisage

[10] Pritchard & Smith (2005). This is far from being the only available philosophical conception of luck. For example, some philosophers define luck in terms of what lies beyond an agent's control: see Nagel (1976), Statman (1991). I cannot hope to address the wider debate about the nature of luck here, but I believe that the Pritchard and Smith account is preferable.
[11] Kripke (1980); Salmon (1982).

modifications of those cases in section 2 which *did* involve luck: modifications in which the element of luck is cancelled, but our unease remains. Roger White (2010) uses this strategy to rule out that Cohen's grad school case has anything do with the fact that Cohen would have to have been lucky to have chosen the right philosophy programme. Perhaps it was highly contingent that Cohen would choose Oxford over Harvard. However, White notes that this is feature is inessential: "Perhaps he couldn't have gone to Harvard. Perhaps he was blacklisted in the United States for his communist sympathies. … [T]hat there was no real risk of his having formed different philosophical opinions seems to do nothing to alleviate the apparent problem Cohen faces." (599)

3.2 White on contingency anxiety

White is generally skeptical that facts about the distal causes of our beliefs have any relevance to whether we should revise our beliefs in cases of contingency anxiety. He notes that such cases are likely to trigger extraneous epistemic worries, which might be doing all the work. One is the problem of Cartesian skepticism. White suggests that one's counterparts in cases of contingency anxiety – those who hold contrary beliefs in light of their differing backgrounds - will tend to activate the familiar skeptical worry that although my judgments appear sensible and correct 'from the inside', they might nonetheless be subject to errors that elude even my best efforts at detection.[12] As supporting evidence, it may be noted that Descartes (1641/1996) himself activates skeptical worries about his rational faculties in roughly this way. He says: "just as I consider that *others sometimes go astray in cases where they think they have the most perfect knowledge*, how do I know that God has not brought it about that I too go wrong every time I add two and three or count the sides of a square, or in some

[12] Cf. Elga (ms.).

even simpler matter, if that is imaginable?" (14. My emphasis.) Cartesian skepticism is not an epistemological problem that we can merely brush aside, but it is also not a problem particularly to do with the contingency of our beliefs on arbitrary background factors.

White also points to a general connection borne by cases of contingency anxiety to the phenomenon of disagreement. As he notes, showing that one's beliefs derive from certain idiosyncrasies of personal background will generally require showing (or rendering salient) that people with different backgrounds hold different beliefs. Disagreement might of itself constitute a reason to revise one's beliefs, and this might be what makes us so nervous, rather than any facts about distal background factors. Thus, with respect to Cohen's case, White thinks that any defeater must derive from Cohen's knowledge that Oxford and Harvard philosophers disagree about the analytic/synthetic distinction. He supports this diagnosis as follows. Suppose, for the sake of argument, that Oxford and Harvard are the only graduate programmes in philosophy. In one scenario, *Correlation*, we suppose that there is disagreement about the analytic/synthetic distinction and it is distributed according to one's choice of graduate programme: the Oxonians believe in it and the Crimsons do not. In *No Correlation*, there is the same level of disagreement with respect to the analytic/synthetic distinction, but it is now randomly distributed throughout the philosophical population. Would an Oxford philosopher who endorsed the analytic/synthetic distinction have any less reason to revise her beliefs in *No Correlation* than in *Correlation*? White sees no clear reason to think that she would. We could imagine taking a similar line on the evidence of cultural variability in epistemic intuitions put forward by Weinberg *et al.* Imagine a counterpart to the *No Correlation* scenario. Suppose Weinberg *et al.* discover that people are far more likely to reject the standard Gettier intuition than philosophers had imagined. However, they uncover no

demographic variables that predict who will deny that subjects in Gettier cases lack knowledge: the Gettier-deniers are randomly distributed across cultural boundaries. Would this be any less worrying? If so, why?

3.3 The Disagreement Hypothesis

My own view is that, for any case of contingency anxiety, any defeater that occurs is due to the epistemic significance of disagreement. I call this the *Disagreement Hypothesis*. While in the spirit of White's paper, he never affirms anything quite so strong. A more important respect in which I differ from White is the following. As we've seen, White treats an emphasis on disagreement as an alternative to thinking that facts about the distal causes of our beliefs enter into our reasons for belief-revision in cases of contingency anxiety. For reasons that I outline in section 5, I do not: the *Disagreement Hypothesis* allows that our awareness of the influence of arbitrary background factors can affect our reasons for revising our beliefs. This, I think, is a virtue of my view. And it allows me to say, for reasons that I will explain, that the *Correlation* scenarios should be more worrying than the *No Correlation* scenarios. There is one final difference between me and White. White is skeptical that there exists any cogent evolutionary debunking argument that could undermine our moral beliefs. I think that at least one evolutionary debunking argument has a hope, and I want to use the *Disagreement Hypothesis* to support this claim.

White's discussion gives some support to the *Disagreement Hypothesis*, but there are two apparently very serious problems that must be overcome if it is to be plausible as a general account of contingency anxiety. I noted these in the introduction: firstly, any element of disagreement can easily seem inessential to cases of contingency anxiety; secondly, placing the emphasis on disagreement seems to rule out any role for knowledge of the relevance of arbitrary background factors, in terms of giving us

reasons to revise our beliefs. I will expand on these issues in the sections that follow, where I show that both problems can be solved.

4. Arbitrarily absent disagreement

4.1 The problem of absent disagreement

Here is the first problem for the *Disagreement Hypothesis*: while many cases of contingency anxiety involve disagreement, some do not, and those that do might seem equally problematic were no disagreement involved.

First off, if we are worried by the thought that our moral outlook is parochial to hominine social life and we count this as a case of contingency anxiety, then not all cases of contingency anxiety appear to turn on disagreement. With respect to those elements of human morality which are supposed to reflect idiosyncratic features of our evolutionary descent, there is relatively little disagreement: after all, these are supposed to be part of human nature. At the least, to the extent that there are people who disagree on these matters, *they* are not the problem we're worrying about when we worry that our moral outlook exhibits phyletic contingency.

It also seems that cases of contingency anxiety which do involve disagreement would be equally problematic if we imagine taking out the element of disagreement, just as White imagined taking out the element of luck in Cohen's grad school case. Cohen (2000) himself makes a move in this direction. He writes:

> suppose I were to discover that I have an identical twin, who was raised not in a communist home but in a politically middle-of-the-road home, and that my twin has the easy tolerance toward limited inequality which I learned to lack. That, I confess, would disturb my confidence in my own uncompromising egalitarianism. … That I am *in fact* twinless should not reduce the challenge to my inherited convictions which is posed by the story I've told. An entirely plausible story could be told about a *hypothetical* disagreeing twin, and it would, or should, be just as challenging as a true story, to those of us who believe what we were brought up to believe. (8-9)

This suggests that disagreement can be stripped out without altering the problem: whether or not there actually exists a counterpart with a different background who disputes one's beliefs doesn't matter.

White (2010) rejects Cohen's claims on this point, saying: "it is hard to see what threat a merely hypothetical disputant poses." (578) Since we are fallible, White notes, the bare possibility of disagreement is always in play, and no more expected conditional on the truth or falsity of what we believe.[13] Actual instances of disagreement do not follow the same pattern. As White says: "It is a *necessary* truth that it is *possible* for someone to disagree with Cohen. …. That someone *actually* disagrees is a contingent matter, and one that is more to be expected given that his arguments are not so good than that they are." (579) In light of White's argument, it may seem implausible that stripping out disagreement should really leave matters untouched, but it is difficult to shake the impression that Cohen is on to something. And consider the following:

Divine Revelation:

Due to certain geological catastrophes, only Western culture exists, but God tells contemporary Western philosophers that had there been a more collectivistic social structure in place here or elsewhere, people would not have been so readily persuaded by Gettier to reject the *JTB Analysis*: they would have had a more holistic cognitive style owing to their social conditions and would not have shared the intuition that Gettier relies on. (If only that supervolcano hadn't exploded in southern China four million years ago, there would be such a society right now, and it would be the most populous on Earth.)

What we learn in this kind of case seems to pose the same kind of problem as the results reported by Weinberg *et al.* – but here there is no one who actually disputes the

[13] Cf. Christensen (2007).

Gettier intuition.

4.2 The Arbitrary Absence Thesis

To get to grips with this issue, I am now going to argue that there are some cases in which the world-boundaries between dissenting verdicts do not matter.[14] I will begin by defending and explaining this claim; I will then show how it solves our problem, and why White's reply to Cohen is irrelevant.

My view is that *arbitrarily absent* disagreement has the same epistemic significance as actual disagreement, all else being equal. This view can be stated more exactly, as follows:

The Arbitrary Absence Thesis:

Necessarily for any S_1, S_2, p: If S_1 believes p and knows that S_2 would believe some contrary of p if not for some condition C, then, if C is arbitrary with respect to S_1's belief that p, S_1 should be as confident of p as S_1 ought to be of p if, all else being equal, S_1 knew that S_2 does believe some contrary of p.

To see why we should endorse this principle, we'll start by considering a case of disagreement that I hope you'll agree involves defeat. I borrow this example from David Christensen (2007):

The Restaurant Case:

You go out to dinner. There are eighteen people in your party: your seventeen companions are experts at mental arithmetic; you are only ordinarily reliable. After dinner, you decide to leave a 20% tip and otherwise split the bill evenly. You all see the bill clearly. Suppose you calculate that your share is $43 each. Everyone else says that your shares come to $45.

[14] For a similar view, see Kelly (2005).

I assume your awareness that the experts disagree with you in this case is defeating. To see why we should accept that arbitrarily absent disagreement is no less worrying, consider what happens when we add some bells-and-whistles to this case. In the first instance, consider:

*The Restaurant Case**:

As in the *Restaurant Case*, except that on discovering that everyone else takes $45 to be your share, you take out your mind-changing ray-gun and zap them into agreeing with you.

Now there is no disagreement, but it doesn't seem that you should be any more confident. Of course, it might be said that there is still disagreement with respect to your belief, albeit *located in the past*. We can overcome this problem by thinking further outside the box. Let's suppose that you have precognitive powers that allow you to predict and prevent undesirable events before they occur. Now, consider the following:

*The Restaurant Case**:*

As in the *Restaurant Case*, except that you first calculate your share of the bill to be $43 and then wait for the maths experts to get to work. You correctly predict that they are all going to arrive at the answer '$45', but you pre-emptively use your mind-changing ray gun and zap them into agreeing with you. They never form the belief in question.

Intuitively your confidence that $43 is the correct answer should be no higher in the *Restaurant Case*** than in the *Restaurant Case* or *Restaurant Case**. Although no one actually disputes your belief in this case, this is due to your own arbitrary intervention; and if a mind-changing ray-gun cannot cancel the epistemic significance of disagreement after the fact, neither can your pre-emptive effort here. This suggests

that arbitrarily absent disagreement has the same epistemic significance as actual disagreement, all else being equal: if a dissenting verdict has been excluded from actuality by something irrelevant, we should proceed as if it had not been so excluded.

This may be thought to tell us something important about the epistemology of disagreement: namely, that insofar as we have reason to change our view in cases of disagreement, disagreement itself is not the issue. There must be some separate, underlying factor that is doing the real work: a factor shared across cases where disagreement is present and where it is absent due to arbitrary factors. How else could it be that we can extract the element of disagreement and leave matters unchanged?[15] I believe, however, that the tendency of actual and arbitrarily absent disagreement to have the same epistemic significance ultimately tells us nothing special about the epistemology of disagreement. Rather, it reflects a more general phenomenon about evidence. In turn, the generality of the phenomenon provides additional confirmation for the *Arbitrary Absence Thesis*.

Consider an analogy. Begin with this case:

Murder!

We are investigating a murder. Lucikly, we have the gun that was used, and Bob's fingerprints are all over it.

Case closed. Now, compare:

*Murder!**

We are investigating a murder. We have the gun, but Bob's fingerprints are not on it. However, we know that Bob's fingerprints would be on the gun if not for the fact that the handle does not absorb fingerprints.

[15] See Kelly (2005: 181-182).

Intuitively, our knowledge in this second case provides no worse evidence for Bob's guilt: arbitrarily absent fingerprints incriminate just as well as actual fingerprints. *Murder!** plausibly stands to *Murder!* as the *Restaurant Case*** to the *Restaurant Case*. However, we are not led to suspect, by comparison of the first pair, that finger-prints are not *really* evidence of guilt: that conclusion is obviously absurd. If anything, the fact that arbitrarily absent fingerprints are strong evidence of guilt is parasitic on the tendency of actual fingerprints to incriminate: if fingerprints were not evidence of guilt, knowing that Bob's fingerprints would have been on the gun would be no evidence of guilt either. By parity, that actual and arbitrarily absent disagreement have the same epistemic significance should not imply that disagreement does not *really* provide grounds for belief-revision. As with fingerprints, we should expect that the epistemic significance of arbitrarily absent disagreement is parasitic on the epistemic significance had by actual disagreement. Thus, if I should suspend judgment in the *Restaurant Case*** this is because I should suspend judgment in the *Restaurant Case* in light of the disagreement occurring there.

4.3 Solving the problem

It should now be easy to see how adopting the *Arbitrary Absence Thesis* solves the first problem I noted for the *Disagreement Hypothesis*. The problem was that cases of contingency anxiety seemed equally problematic if we imagined taking out the element of disagreement. This can be fully accommodated whilst maintaining the *Disagreement Hypothesis*. In cases of contingency anxiety, I, or someone otherwise like myself, would have believed the contrary of what I now believe had some arbitrary background factor been otherwise: if the arbitrary background condition did not hold, you would get a belief that opposes my own. By the *Arbitrary Absence Thesis*, it follows

straightforwardly that whether this dissenting verdict is actualized is irrelevant: since it is known that it would occur but for some arbitrary condition, its epistemic significance is the same whether it is actual or absent. Thus, it is to be expected that we should be no less disturbed by the case described in *Divine Revelation* than by the results reported by Weinberg *et al.*: *Divine Revelation* stands to those results as the *Restaurant Case*** to the *Restaurant Case*. Similarly, with respect to Cohen's claim that discovering he has an identical twin raised to tolerate inequality should be no more worrying than his awareness that he could have had such a twin, we can see that the *Arbitrary Absence Thesis* vindicates the claim, so long as we are allowed the plausible assumption that the conditions that would keep Cohen from having an identical twin raised in this way are entirely arbitrary.

As for White's reply to Cohen, we can see, I think, that White is actually arguing for the entirely plausible, but strictly irrelevant, claim that *merely possible* disagreement can never be a cause for concern over and above our awareness of our own fallibility. White (2010) speaks in terms of "a merely hypothetical disputant" (578). I think the term 'merely' diverts our attention from what is really at issue. In the *Restaurant Case*, if I have my eyes closed in order to concentrate and just arrived at $43 as my answer, I might note that there is some possible world in which all of my dinner companions have arrived at a contrary result. That sort of disagreement is aptly described as 'merely possible'. In the *Restaurant Case***, when I know that my companions would all have arrived at a contrary result, it seems infelicitous to describe this as disagreement that is 'merely possible'. It seems rather more than that. In light of our fallibility, the mere possibility of disagreement is, as White says, never more to be expected conditional on the truth or falsity of what we believe. The same cannot be said about arbitrarily absent disagreement. In the *Restaurant Case***, where I know that my companions would all have arrived at a contrary result, this is not something that I

should have expected in any case: it is more to be expected conditional on the falsity of my view. Once we make clear that we are concerned with the epistemic significance of arbitrarily absent disagreement and *not* merely possible disagreement, White's counterargument should not worry us.

Finally, by considering the analogy with *Murder!/Murder!**, we should not think that the epistemic significance of disagreement is irrelevant in cases of contingency anxiety simply because it is possible for disagreement to be extricated from these cases without reducing our sense of unease. This is entirely compatible with the claim that any defeater arising in cases of contingency anxiety derives from the epistemic significance of disagreement: the epistemic significance of disagreement explains the presence of any defeaters in these stripped-down cases, just as our evidence for Bob's guilt in *Murder!** depends on the evidentiary significance typically had by the discovery of a person's fingerprints on a murder weapon. We just have to say that if disagreement does not occur in some case of contingency anxiety, any defeater that does occur is due to the epistemic significance that such arbitrarily absent disagreement would have had, had it occurred. The epistemic significance of disagreement is still doing the work.

5. The role of arbitrary background factors

5.1 The problem of accounting for the significance of background factors

This section will consider the second difficulty noted in my introduction and show how it can be solved in accordance with the *Disagreement Hypothesis*. Here is the problem. We may worry that if any defeaters arising in cases of contingency anxiety derive from the epistemic significance of disagreement, this rules out that knowledge about the role of arbitrary background factors contributes to our reasons for revising our beliefs. This would be surprising: the relevance of arbitrary background features

seems so central to our unease. It would be remarkable if our anxiety turned out to be entirely misplaced.

To lend additional weight to this issue, consider another example discussed by Cohen (2000). He asks us to imagine that identical twins are separated at birth: one is raised to be a devout Presbyterian, the other to be a devout Roman Catholic. Later in life, they meet and get into a religious argument. The considerations put forward in the argument are familiar to both, and neither is sufficiently impressed by these to alter her opinion. But that is not the end of the story:

> Then each of them realizes that, had she been brought up where her sister was, and vice versa, then it is overwhelmingly likely that (as one of them expresses the realization) *she* would now be Roman Catholic and her sister would now be Presbyterian. That realization might, and, I think, should, make it more difficult for the sisters to sustain their opposed religious convictions. (8)

There seems to be some added kick in that realization. As I'll now argue, the *Disagreement Hypothesis* is compatible with this intuition. It allows that our awareness of the influence of arbitrary background factors can contribute to our reasons for revising our beliefs.

5.2 Explanation and disagreement

When we disagree, it matters why we disagree: the epistemic significance of disagreement is modulated by our explanatory knowledge (or lack thereof) regarding the causes of disagreement. Consider the following:

Macroeconomics:

We disagree about whether immigration decreases the wages of low-skilled workers, but I know that this is because you have not seen the latest paper from the Policy Think Tank.

Here, we disagree, but since I know that you lack some relevant evidence, I have no reason to change my mind. Note that the explanatory significance of the difference in evidence is key. Compare:

*Macroeconomics**:

We disagree about whether immigration decreases the wages of low-skilled workers, and I know that you have not seen the latest paper from the Policy Think Tank. I also know that giving it a look wouldn't change your opinion (you would think the paper riddled with errors and continue to dispute my opinion).

Plausibly, I have less reason to remain steadfast in *Macroeconomics**. And, plausibly, this is because, although you lack the same piece of evidence in both cases, only in the latter does this explain our disagreement. Similarly, consider:

Violin Recital:

We are listening to your daughter play the violin. I judge that she is only mediocre in ability, but you think she is talented. I know that you are generally biased in favour of your daughter.

Plausibly, I should not be moved by your contrary assessment here, because I know that you are generally biased in your daughter's favour. This requires, however, that the bias explains your verdict. To see this, consider:

*Violin Recital**:

We are listening to your daughter play the violin. I judge that she is only mediocre in ability, but you think she is talented. I know that you are generally biased in favour of your daughter. However, I also know that this bias does *not* explain your verdict in this case.

Here, it is much less obvious that I should remain steadfast.

Finally, I want us to consider the following case, due to Jennifer Lackey (2010):

Lunchtime Hallucination:

Estelle, Edwin, and I, have been room-mates for the past eight years. During lunch, I ask Edwin to pass the wine to Estelle, and he replies, 'Estelle isn't here today'. It seems to me clear as day that Estelle is sitting to my right, looking rather befuddled by our exchange. Edwin denies this with equal confidence. Prior to this, neither Edwin nor I had any reason to expect that the other is prone to hallucinations or psychoses.

About this case, Lackey says: "it seems clearly rational for me to continue to believe just as strongly that Estelle is present at the table." (307) I'm not certain that my confidence should remain *entirely* unfazed in this case, but I do feel that I can continue to believe, with reasonable confidence, that Estelle is at the table and Edwin is hallucinating. This may seem to challenge the following principle, proposed by Christensen (2010):

Independence:

"In evaluating the epistemic credentials of another's expressed belief about P, in order to determine how (or whether) to modify my own belief about P, I should do so in a way that doesn't rely on the reasoning behind my initial belief about P." (1-2)

In dismissing Edwin's contrary verdict, it might seem that I have to rely on my own conclusion that Estelle is present: I have no other grounds for supposing that he is hallucinating. However, as Lackey (2010: 309-310) and Christensen (2010: 8-11) note, there is an alternative diagnosis available, consistent with *Independence*. When I

discover that Edwin and I disagree in *Lunchtime Hallucination*, it's clear that one of us is suffering from a major psychological malfunction. Independently of my own verdict, I should be significantly more confident that any such malfunction would be on Edwin's side, because I know a great deal more about myself: for example, I can be more confident in ruling out that *I* have recently suffered a blow to the head. There are many explanations for why we disagree that I can rule out in this way by virtue of my privileged self-knowledge. We can, however, vary the details of *Lunchtime Hallucination* so as to eliminate this feature:

*Lunchtime Hallucination**:

As in *Lunchtime Hallucination*, except that Edwin and I are craniopagus conjoined twins, and so I know him exactly as well as I know myself.

Since Edwin and I are joined at the skull and I lack any privileged self-knowledge, I cannot rely in the same way on my special knowledge of myself to be more confident in ruling out explanatory hypotheses which place the hallucination on my side as opposed to his. I should be significantly more conciliatory in this case.

5.3 Solving the problem

With these points in mind, let's now return to the problem that we noted for the *Disagreement Hypothesis*. We were worried that if any defeater arising in cases of contingency anxiety derives from the epistemic significance of disagreement, this would leave no role for our knowledge of the explanatory relevance of arbitrary background factors. We can now see, I hope, that this worry is misguided. The epistemic significance of disagreement, we know, is modulated by our explanatory knowledge (or lack thereof) regarding the causes of disagreement. Granting the

Disagreement Hypothesis, it should then be unsurprising if the explanatory discoveries that typically enter into cases of contingency anxiety *do* affect our reasons for revising our beliefs.

As an illustration, I want us to go back to a question I posed in section 3. Suppose Weinberg *et al.* had discovered that people are far more likely to reject the standard Gettier intuition than philosophers had imagined, but uncovered no demographic variables that allow them to predict who will deny that subjects in Gettier cases lack knowledge: the Gettier-deniers are randomly distributed throughout the population. I asked: would this be any less worrying?

It would. My reasoning here is by analogy with *Lunchtime Hallucination* vs. *Lunchtime Hallucination**. If there are no demographic variables predicting Gettier-denial, then we don't know why the deniers deny. There would be a number of uneliminated explanatory hypotheses in whose disjunction we might place reasonable confidence: Gettier-deniers might be inattentive; they might lack some degree of conceptual competence or expertise, *etc.* Hypotheses like these have been proposed to account for Weinberg *et al.*'s results: much has been made of the claim that philosophers possess expertise that Gettier-deniers lack.[16] It seems to make a real difference, therefore, that Weinberg *et al.* purport to find not only more disagreement than we might have expected, but disagreement that is predicted and explained by Nisbett's work on the differing cognitive practices fostered by Western and East Asian cultural backgrounds.[17] And it seems to make an added difference that Nisbett traces these differing cognitive styles to something so arbitrary as matters of social identity, economics, and ecology. To the extent that we believe that these factors explain the disagreements in intuition reported by Weinberg *et al.*, we seem to lack independent

[16] See Weinberg et al. (2010), Williamson (2011).
[17] Not everyone agrees, however, that such findings are predicted by Nisbett's work. See Boyd and Nagel (forthcoming).

reasons by which to dismiss the contrary verdicts of Gettier-deniers. And we might reasonably expect that this requires us to be significantly more conciliatory.

More generally, when we find that others disagree on certain fundamental issues - issues on which we are likely to invest great confidence in our own opinions - increasing awareness that the differences between us are traceable to arbitrary background factors will tend to decrease our propensity to downgrade their epistemic credentials, by crowding out explanations which place some obvious cognitive mishap or epistemic demerit on their side.

As this shows, our second problem, like our first, is ultimately no problem at all for the *Disagreement Hypothesis*. We can readily accommodate and explain the phenomenon in question. We should not see the *Disagreement Hypothesis* as an alternative to the view that knowledge of the arbitrary distal factors responsible for our beliefs can affect our reasons for belief-revision in cases of contingency anxiety. Rather, the relevance of these explanatory discoveries can be subsumed under the epistemic significance of disagreement: the former serve to modulate the latter.

6. The *Disagreement Hypothesis* and *Phyletic Contingency*

At this point, the *Disagreement Hypothesis* looks to be in healthy shape. It has seen off what appeared to be its most serious problems, proving itself able to explain and illuminate those phenomena that it seemed least likely to even accommodate.

The *Disagreement Hypothesis* can also readily explain why the paradigmatic cases of contingency anxiety noted in section 2 are more worrying than *Corner Shop* or *My Bayesian Love*. By the *Arbitrary Absence Thesis*, what is known in *Corner Shop* or *My Bayesian Love* about what would have been believed but for some arbitrary background factor should be equivalent in epistemic significance to the discovery of some actual dissenting verdict, like that in *Macroeconomics* or *Violin Recital*, where it

could be known, prior to discovering the fact of disagreement, that one's interlocutor lacks some crucial piece of evidence or the ability to respond appropriately to that evidence. Disagreements of that kind plausibly provide no reasons for us to change our minds. By contrast, the disagreements – actual or arbitrarily absent – discussed in section 2 are of a kind that we plausibly cannot dismiss quite so easily if Christensen's *Independence* principle is correct. I am not here insisting that this principle *is* correct. Indeed, part of what I like about the *Disagreement Hypothesis* is that it may be thought to explain why we are merely uneasy in cases of contingency anxiety, rather than straightforwardly inclined to suspend judgment. We are ambivalent, I think, because we are uncertain and conflicted about the epistemic significance of disagreement.

With these successes in mind, I think we ought to assign significant credence to the *Disagreement Hypothesis*. As a framework for making sense of contingency anxiety, it seems too good to pass on. This, of course, leaves open the question of whether any cases discussed in section 2 *do* involve defeaters. I have merely argued that if they do, this is due to the epistemic significance of disagreement.

Let's return, now, to the issue that really interests us. According to *Phyletic Contingency*, had the conditions for the evolution of moral thought been realized in some distantly related species with a very different form of social organization, we should expect their moral outlook to incorporate certain fundamental differences in moral intuition, appropriate to their form of life. Our concern with the phyletic contingency of our moral outlooks is, I have suggested, a case of contingency anxiety, and any defeater arising in a case of contingency anxiety is due to the epistemic significance of disagreement. Hence, if evidence of phyletic contingency is debunking, this must be due to the epistemic significance of moral disagreement.

The disagreement that concerns us here is not actual; we are, to the best of our knowledge, the only moral animal. However, the fact that these alternate moral

systems are located in other possible worlds is without relevance. They are merely arbitrarily absent, and we know that arbitrarily absent disagreement has the same epistemic significance as actual disagreement, all else being equal. The question, for us, then, is: what would be the epistemic significance of discovering that others disagree with us due to brute differences of intuition, knowing that our differences are ultimately traceable to arbitrary phyletic contingencies?

In light of my answer to the question of what role the influence of arbitrary background factors plays in cases of contingency anxiety, we should conclude that knowledge about distal causes will serve to crowd out the possibility that the contrary verdict is due to discrediting factors that compare unfavourably with those operative in our own background. In effect, there will be nothing to decide between these contrary intuitions except the intuitions themselves: the question of how we should respond will reduce to one of whether we can stick to our own intuition and dismiss theirs as mistaken, or whether a brute clash of intuitions is instead defeating.

7. Conclusion

This chapter is the first in a pair which purports to show that a cogent debunking argument can be constructed by appeal to *Phyletic Contingency*. Here, I've argued that if evidence of phyletic contingency is debunking, this must be decided by reference to the epistemic significance of moral disagreement. I argued for this claim by suggesting that the issue of phyletic contingency is merely a particular instance of a broader epistemic phenomenon, *contingency anxiety*, of which I gave multiple examples in section 2. In keeping with my commitment to look squarely at the underlying epistemological issues relevant to evolutionary debunking arguments, this chapter has sought to address the phenomenon of contingency anxiety in general. In section 3, I showed that certain natural explanations for why cases of contingency anxiety might

involve defeaters fail, and I introduced the *Disagreement Hypothesis* in discussing Roger White's recent work on this issue. I then sought to address two apparent problems for the *Disagreement Hypothesis*, showing that the hypothesis was actually well-placed to provide insightful explanations for the problematic phenomena in question. Overall, I have argued that the *Disagreement Hypothesis* is a very promising framework by which to make sense of contingency anxiety. The *Disagreement Hypothesis* implies that whether we should revise our moral beliefs in light of *Phyletic Contingency* depends on how we should respond to evidence of intuitive disagreement in ethics. The next chapter will argue for a concessive view about the epistemic significance of this kind of disagreement.

6.

Epistemic reasons and persons:

disagreements in moral intuition as defeaters

1. Introduction

Whether evidence of phyletic contingency has the capacity to debunk our moral beliefs turns, I have argued, on the capacity of moral disagreements to provide defeaters. More exactly, given the manner in which our moral beliefs reflect our phylogeny, it turns on the capacity of brute conflicts of intuition to act as defeaters. This chapter will consider the epistemic significance of discovering that we are involved in this sort of disagreement, arguing that the discovery is defeating. As with my discussion in the previous chapter, I will here consider the key epistemological

issues in play at some level of generality, with limited emphasis on their particular application to the case of evolutionary debunking arguments, except as appropriate. As I argued in chapter 2, it is reasonable to believe that a number of moral disagreements in actuality reflect brute conflicts in intuition, and my discussion here is designed to apply equally well to such cases.

In this chapter, I will argue for the following:

The Concessive Principle:

Necessarily, for any persons S_1, S_2, and any moral proposition p: If S_1 believes p with *prima facie* justification on the basis of intuiting p and knows that S_2 believes some contrary of p on the basis of a corresponding contrary intuition, this constitutes a defeater for S_1's continued acceptance of p.

The argument that I give for the *Concessive Principle* is, in the first instance, an argument for the following:

Intuitional Non-Reductionism:

Necessarily, for any persons S_1, S_2, and any moral proposition p: If S_1 knows that S_2 has the intuition p, S_1 has *prima facie* justification for believing p.

The *Concessive Principle* follows from *Intuitional Non-Reductionism* straightforwardly. The latter says that the intuitions of others provide *prima facie* reasons to believe corresponding propositions. The *Concessive Principle* says that the contrary intuitions of others provide *prima facie* reasons to revise our intuitive beliefs. Since the contrary intuitions of others provide *prima facie* reasons to believe contrary propositions, they

provide *prima facie* reasons for us to revise our beliefs.[1] I believe that brute conflicts in intuition are defeating if *and only if* something like *Intuitional Non-Reductionism* is true; I will support this claim in section 2.

The argument I offer for *Intuitional Non-Reductionism* I call the *A Priori Parity Argument*. It represents a modification – a radicalization, if you will - of a line of argument advanced by Alan Gibbard (1990) and Richard Foley (2001), dubbed the *Parity Argument* by Frederick Schmitt (2002). My argument uses *a priori* considerations to ground a form of self-other parity that is grounded empirically under the *Parity Argument*. My argument is also narrowly focused on moral intuitions, for reasons I explain below; the original is more general.

Informally, the *A Priori Parity Argument* runs as follows. We permissibly rely on our own moral intuitions without requiring prior evidence of their reliability. A fundamental self/other asymmetry in epistemology is unacceptable, and so we should extend the same trust to the intuitions of others. Therefore, the intuitions of others also provide *prima facie* reasons for us to believe the propositions they intuit. The informal statement of the argument hides a number of important ambiguities, which I'll resolve as we proceed.

I believe that very similar arguments could be used to establish positions analogous to *Intuitional Non-Reductionism* and the *Concessive Principle* for other fundamental source of non-inferential justification. I've chosen to make the argument specifically about moral intuitions so as to sidestep complications arising from Alison Hills' (2010) recent claim that the epistemology of moral testimony and disagreement is *sui generis*. Hills draws on the widely held conviction that moral testimony is distinctive: there are cases in which we ought not to rely on the testimony of others

[1] This may suggest that the defeaters posited by the *Concessive Principle* are rebutting. The reality is likely to be a little more complicated: see Sinnott-Armstrong (2002: 321-322). Unfortunately, I cannot go into this issue any further here.

regarding moral matters, where trust in their opinion on non-moral matters would otherwise be entirely appropriate.[2] I am not sure whether moral testimony really is distinctive in this way,[3] but I prefer to just avoid the issue.

Here is the plan. In the next section, I provide a bit of dialectical context for my argument, situating it with respect to previous work on the epistemology of moral disagreement. I also offer reasons to think that brute conflicts in intuition are defeating only if something like *Intuitional Non-Reductionism* is true. In section 3, I will argue for a particular interpretation of the thought that we are justified in relying on our own moral intuitions without first having gained evidence of their reliability. As we'll see, there are some *prima facie* plausible interpretations of this informal notion of default self-trust which don't combine with the denial of an epistemic self-other asymmetry to entail *Intuitional Non-Reductionism* or the *Concessive Principle*. I'm going to argue against those views, and for the following:

Intuitional Default Self-Trust

Necessarily, for any person *S*, and any moral proposition *p*: If *S* knows that *S* has/had/will have the intuition that *p*, *S* has *prima facie* justification for believing *p*.

Following that, I will argue for:

No Asymmetry:

Necessarily, for any person *S*: There is no reflexive relation with the kind of epistemic significance that could make it the case that *Intuitional Default Self-Trust* is true but *Intuitional Non-Reductionism* is false.

[2] See Coady (1992: 69-75), Driver (2006), Jones (1999), McGrath (2009), Nickel (2001), Wolff (1998).
[3] See Sliwa (2012) for a recent defence of moral testimony.

Here, I'm going to transpose Parfit's (1984) argument that identity does not matter for survival to show that identity is irrelevant for epistemic normativity in just the way that *No Asymmetry* says it is. I also consider and reject the possibility of attaching epistemic significance to Parfit's Relation R.

2. Dialectical context

As stated, I want to use this section to provide a bit of dialectical context for my argument. I begin, in 2.1, by discussing pertinent work by previous authors addressing the issues that concern us. In 2.2, I will then explain where I am picking up the thread of the debate, and why I believe that others' foundational moral intuitions are capable of defeating our own only if there is some fundamental, *a priori* parity between their intuitions and ours.

2.1 Sidgwick's Principle, Reductionism, and the Parity Argument

Sidgwick's remarks on disagreement have become something of a touchstone in recent discussion, and he appears to have accepted something like the *Concessive Principle*. In a lecture published posthumously, he says:

> in any such conflict [of intuitions] there must be error on one side or the other, or on both. The natural man will often decide unhesitatingly that the error is on the other side. But it is manifest that a philosophic mind cannot do this, unless it can prove independently that the conflicting intuitor has an inferior faculty of envisaging truth in general or this kind of truth; one who cannot do this must reasonably submit to a loss of confidence in any intuition of his own that thus is found to conflict with another's. (1905/2000: 168)

These claims parallel his more widely-known remarks on disagreement in *The Methods of Ethics*:

> if I find any of my judgments, intuitive or inferential, in direct conflict with a judgment of some other mind, there must be error somewhere: and if I have no more reason to suspect error in the other mind than in my own, reflective comparison between the two judgments necessarily reduces me temporarily to a state of neutrality. (1906/1981: 342)

Ralph Wedgwood (2010) attributes to Sidgwick the following principle, which he calls simply *Sidgwick's Principle*:

> If you have a belief about a (first-order) question, and then acquire the (higher-order) information that another thinker disagrees with you about that question, you are rationally required to suspend judgment about that (first-order) question, unless you have *independent* grounds for thinking that the other thinker is less reliable about the question than you are yourself. (224)

Wedgwood rejects *Sidgwick's Principle*. In doing so, he makes an important observation, which applies, *mutatis mutandis*, to my *Concessive Principle*. These principles tell us that we have *prima facie* reason to revise our beliefs when faced with dissenting beliefs and/or intuitions, but they make no requirement that we have any prior evidence suggesting that these dissenting beliefs and/or intuitions are reliable. Of course, evidence that they are *un*reliable, or less reliable than our own, would give us a defeater-defeater. Nonetheless, positive evidence of their reliability is unnecessary to get the initial defeater on the table.

These principles thus attach a kind of 'unearned' epistemic significance to the (contrary) beliefs and/or intuitions of others. They seem to bear an important relation to a position in the epistemology of testimony known as *Non-Reductionism*.[4] Although everyone agrees that the credibility of this position is a core issue in social epistemology, we lack a canonical statement of the view, and a great deal of confusion exists as to whether it's supposed to govern knowledge, justified belief, or warrant.[5] The definition I offer is as follows:

---

[4] See Burge (1993), Coady (1992), McDowell (1994), Reid (1764/1997), Schmitt (1999).

[5] This makes it difficult to say whether Audi (1997b) should be classified as a Non-Reductionist: he treats testimony as an irreducible source of knowledge, but not an irreducible source of justification.

Testimonial Non-Reductionism:

Necessarily, for any S_1, S_2, p: If S_1 knows that S_2 testified that p, S_1 has *prima facie* justification for believing p (even if S_1 has no evidence that S_2's testimony with respect to p-type issues is reliable).

And here is the opposing view:

Testimonial Reductionism:

Necessarily, for anyS_1, S_2, p: If S_1 knows that S_2 testified that p, S_1 has *prima facie* justification for believing p only if S_1 has evidence that S_2's testimony with respect to p-type issues is reliable.

Obviously, neither *Sidgwick's Principle* nor the *Concessive Principle* explicitly concern how we should respond to others' linguistic communications: they are about how we should respond to our knowledge of their mental states. Nonetheless, I think Wedgwood is right that there's a deep connection here. As Jennifer Lackey (2008) notes, there is a general tendency in the epistemology of testimony to conceive of the subject in terms of how we learn from the beliefs and other mental states of our interlocutors, with their utterances treated merely as bridges between minds.[6] Those who accept *Testimonial Non-Reductionism* might therefore be thought to presuppose a yet more basic principle, like the following:

Doxastic Non-Reductionism:

Necessarily, for any S_1, S_2, p: If S_1 knows that S_2 believes p, S_1 has *prima facie* justification for believing p (even if S_1 has no evidence that S_2's beliefs with respect to p-type issues are

[6] This approach is contested by Lackey. See also Moran (2005).

reliable).

It's something like *Doxastic Non-Reductionism* that Wedgwood sees as underlying *Sidgwick's Principle*. And, of course, it's the closely related principle of *Intuitional Non-Reductionism* that I see as underlying the *Concessive Principle*.

Wedgwood rejects *Sidgwick's Principle*, then, because he rejects *Doxastic Non-Reductionism*. In the first instance, he suggests that we have *pro tanto* reasons to reject the principle on grounds of parsimony. If we think of the beliefs of others as providing reasons for us only insofar as we have evidence of their reliability, we can subsume their epistemic significance under more general principles of inductive or Bayesian rationality. By contrast, *Doxastic Non-Reductionism* appears to give a special epistemic status to others' beliefs, which cannot be accounted for in terms of any more general principle. To show that this *pro tanto* preference for theoretical simplicity goes undefeated, Wedgwood attacks prominent arguments that have been taken to support *Doxastic* and/or *Testimonial Non-Reductionism*. He draws on the work of Elizabeth Fricker (1987, 1994, 1995) in criticising the popular argument that our reliance on testimony is too extensive to be supported by first-hand evidence.[7] And he argues against the much less widely discussed *Parity Argument* due to Gibbard and Foley.

I'm going to save discussion of Wedgwood's objection to this argument for the next section: it's actually more germane to my reworking of the *Parity Argument* than to the original. In fact, the *Parity Argument* doesn't really support *Doxastic Non-Reductionism*, *Testimonial Non-Reductionism*, or *Sidgwick's Principle*, for reasons I'll now explain.

I'll focus on Foley's presentation of the argument, which is the most fully-developed. The underlying logic of the argument is encapsulated in Foley's image of

[7] For this argument see, *inter alia*, Burge (1993), Coady (1992), Schmitt (1999).

self-trust radiating outward. It is assumed by Gibbard and Foley that unless we concede defeat to the skeptic, we have to accept that we are justified in relying on our cognitive faculties without prior evidence of their reliability: we are permitted default self-trust. This raises the question of why we should not extend similar trust to others. As Foley (2001) says: "Most of us have prima facie trust in our own faculties even though we cannot give a non-question-begging defense of their reliability. But if so, might not we be rationally compelled to have prima facie trust in others as well?" (101) Gibbard and Foley believe that we are so compelled. The basis of this compulsion is not any *a priori* arbitrariness attaching to the posit of a fundamental self-other asymmetry. (More on that later.) Rather, Gibbard and Foley rely on a host of empirical considerations concerning the relationship of our cognitive faculties to others'. For example, due to our extensive reliance on testimony from early infancy, our most fundamental concepts and assumptions are said to derive from others. Our awareness of these facts, Foley says, pressures us to extend our default self-trust to others: "For, insofar as the opinions of others have shaped our opinions, we would not be reliable unless they were." (101) Trust should be extended on other bases too, including our awareness that others' cognitive faculties have been shaped by similar developmental environments and/or otherwise embody broad commonalities of function. Given self-trust, these commonalities require us to extend similar trust to others, by dint of consistency.

Though the *Parity Argument* is used by Gibbard and Foley to argue that we can be permitted to rely on the word of others without positive evidence of their reliability, the argument does not (and does not purport to) establish full-blown *Testimonial Non-Reductionism. Testimonial Non-Reductionism* expresses a blanket permission to rely on testimony, and *Doxastic Non-Reductionism* represents a similar *carte blanche.* The *Parity Argument* implies that default trust should be extended

outward only in those cases where we have empirical reason to believe that others' cognitive faculties are related to ours in one or more of the ways described above. It wouldn't support automatically trusting the testimony or cognitive faculties of an alien species, therefore. By grounding permissible trust in others in empirical evidence of their similarity to us, this view also rules out *Sidgwick's Principle* and the *Concessive Principle*. As noted by Foley (2001: 108-117), evidence of disagreement is reasonably taken to impede the outward radiation of self-trust, by providing evidence that others' cognitive faculties are *not* relevantly similar to ours, thus undermining the consistency requirement that grounds the outward extension of trust.

2.2. Picking up the thread

Here, then, is where I'm picking up the thread of the debate. I think the *Parity Argument* radically undersells the degree to which we are required to treat others as epistemic equals. In my *A Priori Parity Argument*, I replace the empirical considerations discussed by Gibbard and Foley with an appeal to the *a priori* unacceptability of a self-other asymmetry regarding default trust. Given the permission for default self-trust, this entails trusting the intuitions of others in the same way, regardless of any empirical evidence of shared influence or constitution. The default response to disagreements in intuition, then, is suspension of judgment.

By picking up the thread where I do, I'm tying my discussion of the epistemic significance of brute conflicts in intuition to some quite fundamental questions in epistemology: questions about our trust in others, and about the relevance of identity to epistemic normativity. This might seem like overkill. Granted, if we want to say that contrary intuitions *always* provide *prima facie* reasons for belief-revision, then perhaps these are the sorts of questions we have to take up. But if we want to understand whether evidence of phyletic contingency is debunking or whether the

sorts of brute conflicts of intuition that we find in ordinary life provide defeaters, do we really need to go so far as to debate the merits of the *Concessive Principle* and *Intuitional Non-Reductionism*?

Consider a weaker principle of defeat, with significantly greater *prima facie* plausibility:

The Less Concessive Principle:

Necessarily, for any S_1, S_1, and moral proposition p: If S_1 believes p with *prima facie* justification on the basis of intuiting p, knows that S_2's intuitions are reliable with respect to p-type issues, and knows that S_2 believes some contrary of p on the basis of a corresponding contrary intuition, this constitutes a defeater for S_1.

Couldn't we rely on the *Less Concessive Principle* to infer the presence of defeaters in the cases that interest us? Recall that in Darwin's discussion of the hypothetical morality of honey-bees, he merely seeks to rule out the view according to which "if its intellectual faculties were to become as active and as highly developed as in man, any strictly social animal would acquire *exactly the same* moral sense as ours." (1879/2004: 122. My emphasis.) The phyletic contingency of certain of our moral beliefs might show up simply because we can imagine evolutionary paths that would generate moral systems overlapping with our own to such an extent that we should think them broadly reliable, whilst differing from our own on certain fundamental points. Obviously, this is highly speculative. Less speculative is the idea that the brute conflicts in intuition that we seem to encounter in everyday life involve interlocutors whose moral beliefs agree with our own to such an extent that we should think them broadly reliable on moral issues. So why shouldn't we appeal to the *Less Concessive Principle* in thinking about the significance of these cases?

Here is why I don't see this approach as especially promising. There are two reasons. Firstly, the reasoning set out in the previous paragraph embodies the idea that we should count others' moral beliefs and/or intuitions as broadly reliable because they tend to agree with the moral facts *as we see them*. This gives our own moral outlook some kind of authority with respect to their trustworthiness. Thus, when others' moral beliefs and/or intuitions disagree with our own, this would seem to provide reason for us to downgrade our prior estimate of their reliability and dismiss their view as mistaken, and *not* a reason to suspend judgment.[8] There is a close analogy here with Foley's approach to disagreement: if others get counted as trustworthy on the basis of their similarity to us, points of disagreement counteract the outward extension of trust and permit steadfastness.

Here is a second problem. You could accept the *Less Concessive Principle* but deny that brute conflicts of intuition provide defeaters, even in cases where you know that the person with whom you disagree has moral intuitions or beliefs that are reliable in general. The principle comes into play only when S_1 has evidence that S_2's intuitions on *p-type issues* are reliable. It is not obvious how to type issues. You might think that when *p* is a moral proposition, the relevant type is *moral propositions in general*. It could be said, however, that the force of discovering a brute disagreement in intuition with someone with whom I otherwise agree is to place the matter in dispute in a category of its own. In that case, the *p* about which we disagree would not belong to a type for which I have prior evidence of your reliability, and the *Less Concessive Principle* wouldn't apply. I think this is actually quite a plausible take on the matter. Imagine that our moral intuitions and beliefs are exactly identical, so far as is possible, but for the fact that when it comes to the morality of consensual incest you think it's just wrong, I think it's surely permissible, and we have a brute clash of intuitions. Knowing that two

[8] Cf. Enoch (2010b).

individuals who agree on other issues in ethics nonetheless get locked into a brute disagreement when it comes to the morality of consensual incest, I receive evidence that people's track record on those issues doesn't allow one to predict judging the morality of consensual incest correctly any better than would flipping a coin. I would therefore seem justified in treating the morality of incest as an issue of a kind on which I have poor evidence from which to predict that you'll judge the matter correctly.[9]

In light of these complications, I think it most promising to address the significance of brute conflicts in intuition by thinking in terms of the fundamental epistemological issues that go into the *A Priori Parity Argument*. If we are to view others' foundational moral intuitions as capable of defeating our own, we need a picture on which there is some kind of fundamental parity between us. That is what I intend to supply.

3. Arguing for *Intuitional Default Self-Trust*

In this section, I will be arguing for the particular interpretation of default self-trust represented by *Intuitional Default Self-Trust*. There are a variety of ways in which to interpret the proposal that we are allowed to rely on our cognitive faculties, including our moral intuitions, without prior evidence of their reliability. Not all of these can be combined with the denial of a self-other asymmetry to entail *Intuitional Non-Reductionism*. I'm going to argue that we should prefer *Intuitional Default Self-Trust* over these competitors.

3.1 What basis for self-trust?

To a first approximation, we may understand the notion of default self-trust in moral

[9] Cf. Elga (2007: 492-497).

epistemology as implying:

*Intuitional Default Self-Trust**:

Necessarily, for any *S*, *p*: If *S* has the intuition that *p*, *S* has *prima facie* justification for believing *p*.

One complication about this first approximation should be noted. Since it grants a uniform epistemic significance to moral intuitions and takes no account of characteristically externalist conditions such as reliability, *Intuitional Default Self-Trust** might be thought to commit us to some form of *Epistemic Internalism*. In a sense, I would be fine with that: I think *Internalism* is more plausible than *Externalism*. In general, epistemologists are slightly more likely to accept or lean towards the former.[10] It would be a mistake, however, to suppose that *Intuitional Default Self-Trust** automatically commits us to *Internalism*. A well-known objection to any bald reliabilist theory which counts a belief as *prima facie* justified iff it is the product of a reliable belief-forming process is that it gets the wrong results in a number of non-standard cases, such as those involving clairvoyance[11] or evil demons.[12] To avoid these problems, modifications to the theory have been suggested, which have *Reliabilism* implying that attributions of justification in nonstandard cases should reflect the facts of reliability in standard cases. For example, it has been suggested by Juan Comesaña (2002) that we should adopt *Indexical Reliabilism*, according to which a belief is to be counted as justified even if formed via an unreliable process, so long as it is formed via a process that is reliable in the actual world. Alvin Goldman (1986) suggests that attributions of justification are to follow the facts of reliability in what he calls 'normal worlds':

[10] Bourget & Chalmers (2010).
[11] BonJour (1980).
[12] Cohen (1984).

"beliefs are deemed justified when (roughly) they are caused by processes that are reliable in the world as it is presumed to be." (108) Supposing that we adopt one of these modifications, then if moral intuitions provide *prima facie* justification in the actual world or in the world as it is presumed to be (as I assume they do), *Intuitional Default Self-Trust** follows.

In any case, even assuming *Epistemic Internalism*, the interpretation of default self-trust as *Intuitional Default Self-Trust** doesn't yet get us what we need for the *A Priori Parity Argument* to work. One could, in principle, accept *Intuitional Default Self-Trust**, deny the existence of any self-other asymmetry, and yet reject *Intuitional Non-Reductionism* and the *Concessive Principle*.

We might accept *Intuitional Default Self-Trust** but disagree about its basis. Some may accept what I call *Intuitional Dogmatism*, according to which intuitions of themselves provide basic justification.[13] Others may accept *Intuitional Apriorism*, according to which we have *a priori* default justification for believing our intuitions to be reliable. The parallel here is with corresponding positions in the epistemology of perception. Consider Jim Pryor's (2000) *Perceptual Dogmatism*. Pryor accepts

*Perceptual Default Self-Trust**:

Necessarily, for any *S*, *p*: If it perceptually appears to *S* that *p*, *S* has *prima facie* justification for believing *p*.

Pryor believes that perceptual experiences provide *prima facie* justification simply in virtue of the kind of mental states they are. This justification is *basic*: it does not presuppose any prior justification that one might have to believe any proposition, including the proposition that perception is reliable. As Pryor says: "you have

[13] *E.g.*, Huemer (2005).

justification for believing *p* simply in virtue of having an experience as of *p*. … [T]he *mere having* of an experience as of *p* is enough for your perceptual justification for believing *p* to be in place." (519) Some philosophers are dissatisfied with the idea of basic perceptual justification and insist that if *S* is justified in basing her beliefs on her perceptual states, this requires *S* to have prior justification for believing her perceptual states to be reliable.[14] It may seem that this position leads to skepticism, because there is no way of gaining justification for believing in the reliability of perception without at least some justified perceptual beliefs. To forestall this worry, proponents of this view posit a form of groundless *a priori* justification.[15] Thus, Roger White (2006) says: "we have a kind of default justification for assuming the general reliability of our perceptual faculties. We are entitled to believe that our faculties tend to deliver the truth unless we have some positive reason to doubt this." (552)[16]

Let us call the view just sketched *Perceptual Apriorism* (pun intended). *Perceptual Apriorism* and *Perceptual Dogmatism* agree about the truth of *Perceptual Default Self-Trust**, but they disagree about its basis.[17] According to *Dogmatism*, *Perceptual Default Self-Trust** is true because perceptual appearances provide basic justification. According to *Apriorism*, *Perceptual Default Self-Trust** is true because, necessarily, for any *S*, *p*: *S* has *prima facie* justification for believing that her perceptual states are reliable. With respect to *Intuitional Default Self-Trust**, we can imagine a similar divergence of opinion: *Intuitional Dogmatism* vs. *Intuitional Apriorism.*

Here is why the distinction matters. The *A Priori Parity Argument* says that we

[14] Cohen (2010), White (2006).
[15] Cf. Wright (2004).
[16] While some may balk at the idea of *a priori* justification for contingent propositions, this notion is quite respectable within contemporary epistemology: see, *e.g.*, Hawthorne (2002), Williamson (1986, 1988).
[17] Here I am ignoring the following complication. Some philosophers might think that it is possible for *S* to have the intuition that *p* but have no *prima facie* reason to believe that she has the intuition that *p*. In that case, *Intuitional Dogmatism* would imply that *S* has *prima facie* justification for believing *p*, but *Intuitional Apriorism* would seem to imply that *S* does not. If such cases are possible, *Intuitional Apriorism* is not in fact compatible with *Intuitional Default Self-Trust**. Thanks to John Hawthorne for this point.

are permitted default self-trust and must extend this trust to others, since a fundamental self-other asymmetry in epistemology is untenable. Suppose we accept *Intuitional Apriorism*. Then, we accept *Intuitional Default Self-Trust**, and posit as its basis each person's having *a priori prima facie* justification for supposing that *her* intuitions are reliable. In that case, the reasoning set out in the informal *A Priori Parity Argument* seems cogent: I have *a priori prima facie* justification for believing that *my* moral intuitions are reliable; disallowing any fundamental self-other asymmetry, I should accept a similar principle applying to the intuitions of others. Thus, if we accept *Intuitional Apriorism*, we can resist the argument only by insisting that *my* intuitions have especial epistemic significance *for me*.

Suppose, however, that we accept *Intuitional Dogmatism*. Then, things are not so clear. On this view, intuitions provide basic justification: merely by virtue of having an intuition that *p*, I receive justification for believing *p*. For *Intuitional Dogmatism*, the license for default self-trust concerns the epistemic significance of my *intuitions*, whereas the requirement to trust others expressed by *Intuitional Non-Reductionism* may be said to concern the significance of *evidence that* others have certain intuitions. We do not have to affirm the existence of a self-other asymmetry in epistemology to say that the former should have a kind of significance for us that the latter should not. We simply need to assert a difference between *having an intuition* and *having evidence of an intuition*. It is the whole point of *Intuitional Dogmatism* that intuitions are mental states with a very special epistemic significance: thus, it should not be too surprising if there exists this kind of asymmetry.

Differences of identity might still be said to have derivative epistemic significance on this view. After all, we have different minds. Your intuitions aren't mine, and so can't provide me with basic justification; I can only respond to your intuitions indirectly, by forming beliefs about them. This, it might be said, is why self-

trust cannot radiate outward from *I* to *you*, even if the pronoun *I* has no intrinsic epistemic significance. Wedgwood (2010) offers just this kind of response to the *Parity Argument*. Since he understands the license for default self-trust as a permission "to form moral beliefs directly on the basis of one's current intuitions – even without any additional independent reason for regarding those intuitions as reliable" (240), he denies that there could be any requirement to extend this trust to others, because doing so would be impossible: "It does not seem possible for me currently to form a moral belief *directly* on the basis of *your* moral intuition." (239)

Let's call this stance *Dogmatist Anti-Parity*. As stated, on *Dogmatist Anti-Parity*, we don't attach any real epistemic significance to differences of identity: it's the difference between having an intuition and having evidence of an intuition that counts. Two implications make this clear. Firstly, if I could base my beliefs directly on your intuitions, I would be justified in doing so. (We might imagine science-fictional mind-melding technologies that would permit something like this.) Secondly, although my having an intuition that *p* provides me with *prima facie* reason to believe *p*, my knowing that I *have had* or *will have* an intuition that *p* should provide no such reason, of itself, to believe *p*. Wedgwood (2010: 240-241) fully endorses the view that evidence that I will or have had an intuition that *p* has, all else being equal, the same epistemic significance for me now as evidence that others currently intuit *p*. In other words, I have to rely on evidence for the reliability of my past and future intuitions in order to treat them as reasons to believe corresponding propositions; and in conflicts between present and nonpresent intuitions, I should favour my present intuition, unless I have some special reason not to.

This, as I'll now explain, is the Achilles' heel of *Dogmatist Anti-Parity*. Intuitively, that's *not* how you should respond to cases of intrapersonal disagreement.

3.2 Against Dogmatist Anti-Parity

Let's consider a rather straightforward case:

Dancing Intuitions:

I'm reading about the *Trolley Case* and associated dilemmas when I consider Thomson's *Loop Case* for the first time. It strikes me that you are permitted to turn the trolley in the *Loop Case*: that's my intuition. I go to get a glass of water. I get back to work and re-consider the *Loop Case*. But now it seems to me that the agent may not turn the trolley: that would be wrong.[18] I remember that, minutes ago, I had the opposite intuition. Now I don't know what to think.

In this case, siding with my present intuition seems unjustified: suspending judgment seems the correct response, given that my intuitions dance.

So far, this doesn't clearly rule out *Dogmatist Anti-Parity*. After all, the view doesn't say that you should *never* suspend judgment in cases of intuitive disagreement. It could be said that in *Dancing Intuitions*, my memory of a past intuition has epistemic significance for me because I have justification for supposing that my own past intuitions are reliable. It's just not clear, however, where this justification for believing in the reliability of my past intuitions should come from. Clearly, in *Dancing Intuitions*, I don't gain evidence for the reliability of my moral intuitions in the interim between my first reading of the *Loop Case* and my second. Any justification for the reliability of my past intuitions must be justification that I had upon the first reading for supposing then that my intuitions were reliable. What reason did I have to suppose this? Recall that according to *Intuitional Dogmatism*, we do *not* possess default justification for believing that our moral intuitions are reliable: our intuitions provide justification for believing corresponding propositions *without* our having justification for believing in

[18] For empirical evidence of the instability of *Loop Case* intuitions see Liao et al. (2012).

the reliability of our intuitions. We have to work to gain evidence of reliability. While we can imagine certain ways in which I might have gained this kind of justification, there's no reason to think our intuitions about *Dancing Intuitions* depend on interpolating this kind of detail.

Here is another line of argument by which one might attempt to accommodate *Dancing Intuitions* without attaching any default credibility to our own past intuitions. When I discover that my intuitions on the *Loop Case* dance, it might be thought that this straightforwardly casts doubt on their reliability: just by noting that they disagree, I know that no more than 50% are accurate. Evidence of unreliability might then be taken to undercut my present intuition, which explains why I should suspend judgment in *Dancing Intuitions*.

While *prima facie* plausible, this line of reasoning ultimately falters. Here's the problem. If I were to I intuit that turning the trolley in the *Loop Case* is permissible while knowing that *someone else* had the contrary intuition, then I would also have evidence that my intuition belongs to a reference class whose members are not generally reliable: the reference class comprising mine and my interlocutor's intuitions about the *Loop Case*. The defender of *Dogmatist Anti-Parity* doesn't want to say that there is a defeater whenever there are clashes of intuitions like this: that would be to walk right into the *Concessive Principle*.

How might the interpersonal case just described differ from the intrapersonal case described in *Dancing Intuitions*? It might be suggested that my own intuitions comprise an appropriate reference class because they arise from a common method or procedure, or because I have (more) reason to believe that they do. However, in *Dancing Intuitions*, the defeater that I receive seems very much localized to the one particular case on which my intuitions dance: while I should suspend judgment about the *Loop Case*, it doesn't seem that I am required to do the same with respect to the

original *Trolley Case* or the *Footbridge Case*. It is not at all plausible, however, that I am in possession of a method or belief-forming process whose proprietary domain is the *Loop Case*. If the defeater is due to evidence for the unreliability of the method that I use to decide the *Loop Case*, this should then apply equally to my Trolleyology intuitions across the board; but that seems mistaken. I should suspend judgment only about the *Loop Case*.

To accommodate our intuition about *Dancing Intuitions*, I think we should reject *Dogmatist Anti-Parity* and adopt an interpretation of the notion of default self-trust that is intrapersonal and transtemporal. That's exactly what the first premise in the *A Priori Parity Argument* gives us:

Intuitional Default Self-Trust:

Necessarily, for any person *S*, and any moral proposition *p*: If *S* knows that *S* has/had/will have the intuition that *p*, *S* has *prima facie* justification for believing *p*.

Intuitional Default Self-Trust straightforwardly explains why I should suspend judgment in *Dancing Intuitions*. It also explains why the problem is localized to my intuition about the *Loop Case*: I'm only aware of a dissenting verdict with respect to that case in particular.

I should note that although *Intuitional Default Self-Trust* is to be expected if *Intuitional Apriorism* is true, *Intuitional Dogmatism* doesn't clearly rule out *Intuitional Default Self-Trust*. It is consistent with *Intuitional Dogmatism* to suggest that since intuitions provide basic justification, evidence that certain intuitions will occur or have occurred provides evidence for the intuited propositions.[19]

[19] This would accord with the spirit of the following *Meta-justification Principle*, suggested by White

Ironically, whereas I have been arguing that the posit of an epistemic parity between intrapersonal differences of time and differences of identity marks the Achilles' heel of *Dogmatist Anti-Parity*, Wedgwood regards this aspect of his view as a strength, and as getting the cases right. Speaking of future intuitions, he writes:

> it does not seem so strange to me that, even if one receives the information that one will have a certain intuition in the future, one's response to this information should be guided by one's rational assessment of whether one's intuitions can be expected to become more or less reliable in future. There is also no need for this assessment to be independent of one's current intuitions. Indeed, in some cases, the very information that I will come to have a certain intuition in future gives me reason to think that my moral sensibilities will deteriorate. For example, suppose I receive the information that, in the future, I will have the intuition that it is an admirable form of tough-mindedness for the police to have a policy of torturing those whom they suspect of serious crimes. This information would seem to me all by itself to make it rational for me to think that my moral intuitions will probably be less reliable in future than they are now. (241)

In spite of everything said so far, Wedgwood's claims about his future torture intuition seem *prima facie* convincing. Wedgwood takes this to support the idea that I should be partial to my own present intuition in just the way that I ought to be partial to my own intuition in disagreeing with others, flouting *Sidgwick's Principle*. However, I can readily accommodate the view that Wedgwood should be confident that his future intuition will be mistaken without back-tracking on anything I've said, since Wedgwood *does* have independent reason to support his confidence on that point.

My reasoning will be familiar from section 5.2 in the previous chapter. In Wedgwood's case, I'm told that my future intuition will favour torture, and nothing more. This leaves open a host of questions about what will happen to me. I don't have to rely on the correctness of my present view in order to remain reasonably steadfast in that case. I can rely on the knowledge that, since one of us must be badly mistaken,

---

(2006: 538): "If S is justified in believing that he will be justified in believing p, then he is already justified in believing p." (However, White relies on this principle in arguing against *Perceptual Dogmatism*.) See also Van Fraassen (1984) and Briggs (2009).

something must have gone wrong somewhere. I can combine this knowledge with my more detailed knowledge of my present self: this should lead me to place greater confidence in the hypothesis that I'll be subject to some relevant epistemic fault in future, because my privileged knowledge about my present self allows me to rule out obvious sources of bias and poor judgment in me as I am now. In *Dancing Intuitions*, by contrast, there are no similar considerations that could support reliance on my present intuition because I know a great deal about what happens in the intervening period between the contrary intuitions and it is utterly arbitrary: I went to get a glass of water. Here, I should suspend judgment.

4. *Arguing for No Asymmetry*

I will assume that I have done enough to support the first premise in the *A Priori Parity Argument*. To seal the deal, I need to argue for the second. Section 4.1 re-states the premise and considers why some might be attracted to its denial. Section 4.2 argues that *No Asymmetry* should be accepted in any case, relying on intuitions about examples modified from those discussed by Parfit (1984) in his work on personal identity and survival.

4.1 Self-other asymmetry in epistemology and ethics

The second premise, we recall, is as follows:

No Asymmetry:

Necessarily, for any person *S*: There is no reflexive relation with the kind of epistemic significance that could make it the case that *Intuitional Default Self-Trust* is true but *Intuitional Non-Reductionism* is false.

There's an element of shadowboxing in my argument in this section, as I don't know of anyone who has explicitly endorsed a fundamental self-other asymmetry of the kind that *No Asymmetry* denies. Nor can I imagine any clear reason for doing so - with one exception. I can imagine that people might be inclined to deny *No Asymmetry* if they thought that doing so was necessary to rule out the concessive implications of the *Concessive Principle*. Those implications are likely to be revisionary of common epistemic practice. As Huemer (2011) says: "very few people place as much confidence in another person's intuitions as they would place in their own." (18) While I expect that much of this can be defended by reference to asymmetries in self-knowledge, I doubt that this is sufficient to legitimate common practice wholesale. The *Concessive Principle* is likely to generate resistance from some people, and they might rally to the posit of a self-other asymmetry.

To bolster their case, they might appeal to certain aspects of practical normativity as companions in guilt. Unlike Godwin, most people believe that there is some magic in the pronoun 'my'. For example, they believe that each person has especial reason to be concerned with *her own* welfare, as a matter of prudential rationality. They may also believe that "each of us is especially responsible for what *he* does, rather than for what other people do." (Williams 1973: 99) In the same vein, they might attach value to virtues of integrity or authenticity, which seem to involve an essentially reflexive element: being true to *oneself*. My own strategy, however, is to play the analogy with practical normativity *against* the posit of a self-other asymmetry in epistemology.

4.2 No Asymmetry, personal identity, and fission

In arguing for *No Asymmetry*, I'll draw on cases first used by Parfit (1984) in questioning the practical significance of the boundary between self and other in his

discussion of *identity fission.* In doing so, I'm going to be assuming the *Psychological Criterion* of personal identity. This assumption is designed purely to focus our discussion, by fixing in mind a plausible criterion for identity over time. The argument can be made on any account of personal identity that allows for fission. According to Harold Noonan (2003), "any criterion of personal identity *at all*" (127) allows for fission.

4.2.1 Fission, metaphysics, and prudential concern

Fission occurs when the relations that normally support identity over time hold between *one* past or present individual and *two* future individuals. Thus, if we suppose that identity over time is normally supported by psychological continuity, a case of fission is one in which two later individuals are psychologically continuous with one present individual. Parfit describes such a case, as follows:

> *My Division.* My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people believes that he is me, seems to remember my life, has my character, and is in every other way psychologically continuous with me. And he has a body that is very like mine. (254-255)

Fission cases are important for our understanding of the metaphysics of identity. If we wish to maintain the *Psychological Criterion* of personal identity in some form whilst respecting the intuition that there are two distinct individuals existing at the end of *My Division*, we cannot say that identity over time consists in psychological continuity. Instead, we must say that identity consists in *non-branching* continuity.

Fission cases also tell us something important about the ethical significance of identity. As noted, we tend to think that some degree of prudential concern is rational. Consider an adult person, Sally. Amongst all those individuals who exist in the future,

we are inclined to think that Sally should be particularly concerned with the wellbeing of her future self: she has a 'stake' in that person's welfare that she does not have in the welfare of others. We call this special concern that Sally ought to have for her future self, *prudential concern*. We might believe that, if prudential concern is rational, then it is the relation of identity that grounds the rationality of prudential concern: Sally has a special stake in the wellbeing of her future self because her future self is her future *self*. An alternative proposal, favoured by Parfit, is that it is psychological continuity and/or connectedness - *Relation R* – that grounds prudential concern.

When we consider cases of fission, it seems most plausible that Parfit is right to favour Relation R as the ground of prudential concern. Suppose that identity over time consists in non-branching psychological continuity. Then, were one of the transplants in *My Division* to fail, I would continue to exist; it is only when both transplants are successful that I cease to exist. Two successful transplants don't seem to add up to something as bad as death. Everything that is ordinarily necessary and sufficient for survival and prudential concern seems to be there, merely duplicated. As Parfit says: "Dying is one, dividing is another. To regard these as the same is to confuse two with zero." (262) This suggests that Relation R must be what gives us a special prudential stake in the future. This has important implications, since Relation R, unlike identity, holds by degrees, and may hold to some extent amongst distinct persons – even those outside science-fictional scenarios like *My Division*. As Parfit says: "There is still a difference between my life and the lives of other people. But the difference is less." (280)

4.2.2 Fission and the epistemic significance of identity

We can use fission cases to test our intuitions about the significance of identity in epistemology. We are supposing that *Intuitional Default Self-Trust* is true. Let's then

consider the following case:

Possible Bisection:

I have been fatally injured. I know that at least one half of my brain will be transplanted, and there will be psychological continuity. I know that someone with this part of my brain will later have the intuition that it is permissible to turn the trolley in the *Loop Case*. I am presently unsure, however, whether there will also be a second transplantation.

Suppose we believe that identity matters in epistemology in the way that *No Asymmetry* is supposed to rule out. In *Possible Bisection*, we should then believe that I have reason to believe that turning the trolley is permissible in the *Loop Case* if I gain evidence that there will *not* be a successful second transplantation of the same kind. By contrast, evidence that a second, successful transplantation will occur would undercut my reasons for believing that turning the trolley is permissible, because then the intuition wouldn't be mine. This holds even if I should expect the person at the end of the second transplant also to have the intuition that turning the trolley is permissible. This implication is bizarre. It just does not seem that facts about an additional transplantation could undercut my reasons for believing that turning the trolley is permissible in this way.

Even if we accept this verdict, we haven't yet established *No Asymmetry*. *No Asymmetry* denies that there is any reflexive relation such that *Intuitional Default Self-Trust* could be true but *Intuitional Non-Reductionism* false. If we give up on the idea that the identity relation has any deep significance, we might be inclined to attribute this kind of significance to some other reflexive relation. In this respect, Relation R might seem a natural fall-back.

This inclination should be resisted. To see why, we first need to get a little

clearer on how to understand Relation R. Following Parfit, we say that *connectedness* consists in the holding of *direct psychological connections*; *strong connectedness* involves a sufficiently large number of such connections;[20] and *continuity* consists in overlapping chains of strong connectedness. Parfit offers three possible interpretations of when a direct psychological connection obtains. On what he calls the *widest version*, we allow any causal connection amongst mental states to subserve a direct connection: thus, we say that there is a direct connection between S_1 and S_2 if, for example, S_2 has an experience e at t_1, S_1 has a memory as of e at t_2, and the former fact figures in the causal explanation of the latter.[21] Alternatively, we may want to set restrictions on the nature of the causal connection between S_2's experience of e and S_1's memory as of e. Following what Parfit calls the *narrow version*, we might say that the apparent memory must depend upon the earlier experience *in the normal way*. A middle position between the narrow and the widest version is the *wide version*, which maintains that a direct connection holds whenever the causal relation between two mental states is of a kind that reliably subserves such a connection.

I think the narrow understanding of what a psychological connection consists in is least plausible. We should despair of the possibility of finding a description of the 'normal' connection between an experience and the memory of it that could encompass the full range of possible minds, biological and artificial, whilst enforcing any genuine restrictions. Human memory works in a particular way and exhibits certain characteristic properties, such as 'chunking'. Other memory systems might work very differently, whilst being memory systems nonetheless.[22]

If we assume that the wide or widest understanding of psychological

[20] According to Parfit: "we can claim that there is enough connectedness if the number of direct connections, over any day, is *at least half* the number that hold, over every day, in the lives of nearly every actual person." (206)

[21] Memory is not the only direct psychological connection in Parfit's exposition of the *Psychological Criterion*, but it is most useful for discussing the difference between the narrow, wide, and widest versions.

[22] See Clark (2005) and Sprevak (2009).

connectedness is right, it can then be shown that granting epistemic significance to Relation R would have absurd consequences. Testimony is, I assume, a reliable means of getting a belief from one brain to another: there must therefore be a direct connection between myself and another whenever I believe a sincere assertion, and so some degree of R-Relatedness. Imagine that I am listening to a speaker making various assertions. I know that the assertions made by the speaker correspond to the contents of her intuitions. Suppose that I have no reason to believe any of her assertions: accepting any of her assertions would be unjustified. However, I must extend default trust to the intuitions of others, to the extent that I have reason to believe that I am R-Related to them. This implies that if I unjustifiably begin to accept her assertions, I can gain increasing justification for accepting further assertions; I can pull myself towards justified trust by the bootstraps of my own gullibility. This seems unacceptable. Therefore, we should reject the idea that Relation R has this kind of epistemic significance.

Relation R was the natural fallback once we decided that identity couldn't have any fundamental epistemic significance. Now Relation R is out of the picture too. There is, it seems to me, no natural fallback waiting to take over once Relation R is dismissed. It seems most plausible that there is no reflexive relation with the kind of epistemic significance that could make it the case that *Intuitional Default Self-Trust* is true but *Intuitional Non-Reductionism* is false. That, of course, is just what *No Asymmetry* says. So *No Asymmetry* is true.

5. Conclusion

I have now argued for both *Intuitional Default Self-Trust* and *No Asymmetry*. Together, these imply *Intuitional Non-Reductionism*, and *Intuitional Non-Reductionism* entails the *Concessive Principle*. Thus, we should accept that brute conflicts in intuition provide

defeaters.

Having established in chapters 3 and 4 that the most prominent arguments that seek to derive debunking implications from *Functional Truth-Irrelevance* without appeal to *Phyletic Contingency* fail, this completes my argument that we can build a cogent debunking argument by appeal to *Phyletic Contingency*. In the previous chapter, I argued that if evidence of phyletic contingency is debunking, this must be decided by reference to the epistemic significance of moral disagreement. Here, I hope to have shown that the epistemology of moral disagreement decides in favour of the view that awareness of phyletic contingency is defeating. Thus, evolutionary explanations have the capacity to defeat our moral beliefs insofar as they reveal that certain elements of our moral outlook reflect our parochial status as hominine mammals.

7.

Conclusion and directions for future research

1. Introduction

In this final chapter, I'm going to briefly summarize the key points of my argument and outline some important questions that arise in light of my conclusions. These questions have to do with the scope of evolutionary debunking arguments and their meta-ethical presuppositions. I won't try to resolve these questions, but simply indicate the kind of issues that are likely to arise in the attempt. Section 2 will provide the summary, sections 3 will address questions about meta-ethical presuppositions, and section 4 will take up the issue of scope.

2. A brief summary of my argument

This section offers a précis of the key points in my argument, designed to refresh our memory and ensure we understand how questions about scope and meta-ethical presuppositions are likely to be affected by my conclusions.

Let's begin at the beginning. As set out in my introduction, the primary question I've sought to address in this essay is whether evolutionary explanations can debunk our moral beliefs by supplying defeaters for at least some of our moral judgments. I've addressed this question by exploring what evolutionary explanations must be like in order to be debunking, whether we have evidence for explanations of that type, and, most importantly, what kind of epistemic principles are required to establish the presence of defeaters given such evidence. Most contemporary discussion in this area, I've noted, is centred on the question of whether debunking implications follow from *Functional Truth-Irrelevance*, which contrasts with the emphasis on *Phyletic Contingency* characteristic of Victorian concerns about the debunking power of evolutionary explanations. I've positioned myself as an originalist. By considering the most prominent arguments in the literature today, I've offered reasons to think that debunking arguments centred on *Functional Truth-Irrelevance* don't work, and that a successful debunking argument requires us to appeal to *Phyletic Contingency*.

Here is how I proceeded. To assure ourselves that we're not merely tilting at windmills, I first examined contemporary scientific evidence about the evolutionary origins of morality to establish that we have sufficient reason to accept both *Functional Truth-Irrelevance* and *Phyletic Contingency*. I argued that we have such evidence. I then considered whether we could identify plausible epistemic principles that would allow us to infer the presence of defeaters from evidence of this kind. In chapters 3 and 4, I considered three prominent lines of argument that seek to derive debunking implications from *Functional Truth-Irrelevance*: the first appealed to *Ockham's Razor*, the second to considerations of Nozickian sensitivity, and the third to the *Coincidence*

Problem. In each case, I identified significant flaws that led me to believe that none of these arguments should convince us to alter our moral beliefs (including our meta-ethical beliefs). In chapters 5 and 6, I then sought to construct a cogent debunking argument rooted in *Phyletic Contingency*. Firstly, I asked us to see our concern with the phyletic contingency of our moral beliefs as an instance of the broader phenomenon of contingency anxiety, and I argued for the *Disagreement Hypothesis*, according to which the presence (or absence) of defeaters in cases of contingency anxiety depends on the epistemology of disagreement. Appealing to the *Arbitrary Absence Thesis*, I inferred that whether evidence of phyletic contingency has the capacity to debunk our moral beliefs turns on the capacity of moral disagreements to generate defeaters: more exactly, on the capacity of brute conflicts of intuition to act as defeaters. In chapter 6, I used the *A Priori Parity* argument to support the *Concessive Principle*, according to which brute conflicts of intuition do supply defeaters. I argued that the *Concessive Principle* follows from the default permission to trust our own intuitions and the inappropriateness of a fundamental self/other asymmetry in epistemology, relying on cases such as *Dancing Intuitions* and *Possible Bisection* to guide our judgment on this issue.

From this I conclude that evolutionary explanations can debunk our moral beliefs insofar as they provide evidence that our moral intuitions reflect arbitrary contingencies of our hominine ancestry. In addition, my objections to the debunking arguments most prominent in the literature today lead me to believe that we lack adequate reason to suppose that evolutionary explanations are debunking *apart from* their ability to provide evidence that our moral outlook is parochial to our phylogeny.

We are now going to consider how these conclusions impact our understanding of the relevance of meta-ethical assumptions to the debunking power of evolutionary explanations and the scope of evolutionary debunking arguments.

3. Meta-ethical presuppositions?

In my introduction, I noted that one important question in contemporary discussion concerns whether evolutionary debunking arguments require particular meta-ethical presuppositions. Some philosophers think that evolutionary considerations can establish the presence of defeaters for our moral beliefs only if the moral facts are assumed to be constitutively independent of our evaluative attitudes, whereas others dispute this. I suggested that whether our meta-ethical views are relevant to the debunking power of evolutionary explanations turns on why we should think of evolutionary explanations as debunking. What, then, about my attempt to derive debunking implications from *Phyletic Contingency*? Does this require any particular meta-ethical presuppositions, over and above the assumption that our moral commitments are evaluable in terms of epistemic justification?

The argument I put forward in chapters 5 and 6 relied primarily on claims about the epistemic significance of disagreement, such as the *Arbitrary Absence Thesis* and the *Concessive Principle*. To determine whether our meta-ethical view could make a difference to the acceptability of my argument, it seems we should consider which meta-ethical views have relevance for the epistemology of moral disagreement. Unfortunately, so far as I'm aware, this question has received scarcely any attention within contemporary meta-ethics. In the remainder of this section, I'll offer a few brief comments that might serve as leads for further investigation.

Firstly, I find it difficult to see that a preference for or against a view of moral facts as mind-independent could make a difference to the epistemology of disagreement. If viewing the moral facts as constructed entails some alternative epistemology for ethical disagreement that would allow us to reject the *Arbitrary Absence Thesis* or the *Concessive Principle*, this is far from obvious. All else being equal,

the norms for disagreement don't appear to vary according to whether the matter in dispute is mind-independent or constitutively dependent on our attitudes and/or practices. Consider an example drawn from outside ethics. How we should react to disagreement about the causes of the End Permian mass extinction and the current exchange rate for yen and sterling might vary according to what we think about the evidence available to our interlocutor or her track record in assessing evidence of that kind, but it does not seem to make any difference of itself that the exchange rate is a matter of social convention whereas the causes of the End Permian mass extinction are not. Absent some reason to think that we shouldn't extrapolate from this case to the issue of moral mind-in/dependence, we should think that the latter is irrelevant to the epistemology of moral disagreement.

There might be other meta-ethical positions with important implications for the epistemology of moral disagreement.[1] The most obvious candidate, I believe, is *Moral Relativism* of the kind set out in recent work by Max Kölbel (2002, 2004a, 2004b).[2] The key feature of this view is that it allows for *faultless disagreement.* A faultless disagreement occurs when two individuals believe contradictory propositions without either being mistaken. The possibility of faultless disagreement is typically illustrated by considering disputes in taste, such as a disagreement about whether rollercoasters are fun or whether Chaplin is funny: many are inclined to think there is

[1] Schafer (2012) claims that *Quasi-Realism* permits steadfastness in cases of moral disagreement where *Meta-Ethical Realism* would require consilience, but does not explain why. Blackburn (1996) argues in favour of a view on which steadfastness is typically permissible in cases of deep disagreement, but offers no clear explanation for the relation of this view to *Quasi-Realism*; like the realist Enoch (2010b), he appeals primarily to the unavoidability of the first-person perspective in adjudicating disagreement. Gibbard (2003: 284-287) appears to suggest that *Quasi-Realism* does not affect the epistemology of disagreement, and has implications for moral epistemology only in relation to the issue of *deep vindication* (see Gibbard 2003: 251-267).

[2] For a similar view, see Brogaard (2008, 2012). Kölbel describes this view as *Genuine Moral Relativism* in order to contrast it with the *Indexical Relativism* favoured by Dreier (1990), Harman (1996), and Wong (1984, 2006). The latter is relativistic only in that it allows moral sentences to express different propositions according to the context of utterance. MacFarlane (2005, 2007) defends a yet more radical relativistic view.

no unitary standard of truth in these cases.[3] Recent experimental evidence gathered by Hagop Sarkissian and colleagues (2011) suggests that people are inclined to attribute faultless moral disagreement in disagreements involving individuals from very different cultures and/or ways of life.

Moral Relativism of the sort outlined by Kölbel allows that neither party need be mistaken in a disagreement of this kind because propositions should be evaluated as true/false relative to the perspective of a speaker, much as we are accustomed to the idea that the truth of a proposition is relative to a possible world. In cases of faultless disagreement, each party to the dispute can then be counted as correct, provided we allow that truth is here relativized to different perspectives inhabited by the different speakers and that each believes a proposition that is true according to her perspective. Where a disagreement is faultless, it seems natural to deny that either party has any reason to alter her belief, because neither party need be in error. Kölbel (2004a) says: "even though they disagree, giving up the belief in question would be an improvement for neither of them." (53) Thus, if we are inclined to accept *Moral Relativism* we may be inclined to reject the *Concessive Principle* and, with it, my debunking argument. The plausibility of *Moral Relativism* could thus be thought an important question in deciding whether my argument should ultimately lead us to revise our moral outlook.

We should take care, however, in inferring that my argument has *presupposed* the falsity of *Moral Relativism*, as opposed to merely being an indirect argument against that view. This point applies more generally. It is one thing to identify a meta-ethical position according to which the epistemology of disagreement differs from the account for which I've argued, and quite another to show that my argument has assumed prior justification for ruling out that meta-ethical position. Absent evidence of that assumption, we should think that my argument simply doubles as an argument

[3] Whether such cases *are* best interpreted as cases of faultless disagreement is, of course, contentious. See Cappelen & Hawthorne (2009) for a notable argument to the contrary.

against the meta-ethical position in question.

Looking back over my argument in chapters 5 and 6, it is difficult, I think, to see any point at which any particular meta-ethical presupposition was made. In arguing for the *Arbitrary Absence Thesis*, *Intuitional Default Self-Trust*, and *No Asymmetry*, I proceeded by consulting our intuitions about cases like *Dancing Intuitions* and *Possible Bisection* and identifying epistemic principles that best accommodate those intuitions. If any particular meta-ethical view was assumed, the assumption was highly implicit and correspondingly unobvious. I'm inclined to believe that no questions have been begged against any minimally plausible meta-ethical view and that any view which contradicts my conclusion is simply called into question on that basis. However, more careful consideration of this issue would be desirable.

4. Scope

Another issue that has gained prominence in contemporary discussion has to do with the scope of evolutionary debunking arguments. As noted, some philosophers believe that morality as a whole is called into question if we accept that evolutionary explanations can debunk our moral beliefs, whereas others believe that evolutionary debunking arguments can be deployed more selectively. How does the view I've put forward in this essay interact with this debate?

As it's usually conducted, the debate turns on whether certain of our moral beliefs might be suitably independent of the influence of natural selection, with plausible candidates identified by seeking out beliefs and/or corresponding behaviours that appear unlikely to have been adaptive. Moral beliefs such as *Utilitarianism* or the *Golden Rule*, which require a strict form of impartiality and/or universal benevolence, have been held up as most promising in this respect by philosophers including Peter Singer (2005) and Derek Parfit (2011). In reply, critics like Guy Kahane (2011) suggest

that we can explain these beliefs as reasoned extensions of the more restricted forms of benevolence and impartiality that we expect natural selection to favour, placing these beliefs within the scope of discrediting evolutionary explanations after all.

This debate presupposes that the problematic feature of evolutionary explanations has to do with the existence of functional explanations which make no reference to the accuracy of beliefs of the kind that they explain. On the view I favour, a moral belief might be an innate adaptation or arise quite directly via the reasoned extension of such a belief, and yet evidence of this kind would not be counted as debunking unless accompanied by corresponding evidence that the belief reflects arbitrary facets of our status as hominine mammals. Even supposing that Kahane is right about the distal causes of beliefs involving unrestricted partiality and/or universal benevolence, it's unclear that a verdict of phyletic contingency follows alongside. Suppose that beliefs of that kind do result from the rational revision of the more limited sympathies favoured by natural selection via the elimination of arbitrary distinctions and the pursuit of greater coherence and simplicity. We might expect that any moral system must incorporate some degree of benevolence and/or impartiality in order to count as a moral system.[4] If rational reflection necessarily tends toward 'expanding the circle', we might then suppose that rational revision of the initial fund will in every case converge, at the limit, on beliefs prescribing universal benevolence and impartiality. Consequently, these beliefs will exhibit no element of phyletic contingency. These are obviously highly speculative remarks, but they give some indication of how the debate about scope is refashioned in light of my conclusions.

As this makes clear, I think much of the contemporary debate about the scope of evolutionary debunking arguments has been misdirected. The question is not how much of our moral outlook lies within the explanatory purview of natural selection,

[4] For this suggestion see, *inter alia*, Foot (1958), Jackson (1998: 129-138), Smith (1994: 40).

but how much of our moral outlook is contingent on our phylogeny. In chapter 2, I argued that at least some of our moral outlook is likely to be parochial to our hominine ancestry. However, I did not say how much of our moral outlook this covers. Nor, admittedly, was I altogether specific about which particular moral beliefs reflect phyletic contingency, though I did identify proposals by prominent evolutionary scientists, including our beliefs about obligations amongst kin and property rights. To these I think we should add those of our beliefs which most likely reflect our nature as innately hierarchical animals in which the active suppression of hierarchy has been necessary for success in group competition. These may be thought to include beliefs which attribute intrinsic value to certain forms of distributive equality, autonomy, and participatory government.

In thinking further about the issue of scope, a fruitful strategy may be to proceed by attempting to identify plausible constraints that might delimit the potential for moral diversity across taxa. Earlier I mentioned the suggestion that any moral system must prescribe some degree of benevolence and/or impartiality in order to count as a moral system. If that is true, then human morality exhibits no phyletic contingency in this respect. What might ground the existence of a constraint of this kind? Here I'll briefly consider two proposals, the first drawn from philosophical semantics, the second from the philosophy of biology.

Firstly, according to a view held by a number of philosophers, disagreement can never be total: disagreement is intelligible only against the backdrop of extensive agreement, which is required to fix a common subject-matter. This view is particularly associated with Donald Davidson's (1969, 1973, 1974) work on truth and interpretation, where the necessity of agreement for disagreement forms the basis of the famous claim that radical interpretation should be guided by the *Principle of Charity*. Davidson and a number of philosophers working under his influence suggest

that the scope of potential moral disagreement is sharply limited: there are important constraints on the extent to which we can credit others as holding moral beliefs which are false by our lights if we are to credit them as holding moral beliefs at all.[5] This view is not particular to Davidson's semantic programme. We find roughly the same view in Frank Jackson's (1998) well-known attempt to apply the Canberra Plan to ethical discourse. In line with the general contours of the Canberra Plan, Jackson supposes that moral terms are defined by their role within a network of platitudes, which together constitute 'folk morality'. Sufficiently many of these platitudes must be accepted, Jackson claims, if we are to count as sharing a moral language and capable of disagreeing: "Genuine moral disagreement, as opposed to mere talking past one another, requires a background of shared moral opinion to fix a common, or near enough common, set of meanings for our moral terms." (132)

Another possibility is that certain biological factors might delimit the potential for moral diversity across taxa. Here I'll note one particular development of this line of thought that I find especially promising.

Morality is a biological trait characteristic of human social groups but not limited, of necessity, to our slender branch on the tree of life. Many traits can be found across different and distantly related taxa, and traits so distributed do exhibit significant variability. For example, eyes are abundant throughout the animal kingdom and are highly diverse.[6] Given the extent of real and imaginable variability, it's natural to ask on what basis the various biological structures we classify as representatives of a common trait manage to constitute a unitary category. According to one widely

[5] See Cooper (1978), Davidson (1995), Hurley (1985, 1989), Stout (1988). In addition, Davidson's views on mutual intelligibility are often taken to imply that any disagreement is in principle capable of resolution on the basis of shared criteria of correctness. Hence, Davidson's semantic programme is thought to rule out the possibility of rationally irresolvable disagreement. See Davidson (1995), Hurley (1984, 1989), Tersman (2006: 36-40). For critical discussion see Lillehammer (2007).
[6] See Land & Nilsson (2012).

accepted answer to this question, they are unified by shared function.[7] An eye is not an eye in virtue of any key morphological characteristics, but in virtue of having the function to allow vision. The eyes of alien species might be as weird and wonderful as we can imagine, but they count as eyes iff they have the same function as the globules located in the front of our skulls.

If we accept a view of morality as a biological phenomenon, we are then in a position to derive one important constraint on the extent of possible moral diversity across taxa: any system of moral norms must, to count as such, have the same function as the moral norms characteristic of human societies. What is that function? The nature of functions is a contested issue within the philosophy of biology. Within the context of evolutionary biology, the most widely accepted account, the *Selected-Effect Account*, treats the function of a trait as a matter of its selection history: the function of a trait is to produce the effect whose past production explains selection for the trait.[8] Assuming this holds true, the adaptationist hypothesis for which I argued in chapter 1 entails that the function of morality is to suppress competition and promote cooperation in such a way as to promote group fitness. Accepting that traits are individuated by their functions, it follows that any other system of moral norms that might be expected to emerge from the evolutionary process must share this function. This implication is not closely wedded to the *Selected-Effect Account*, I think. Accounts that identify functions with the current causal roles of traits[9] will arguably support similar conclusions, as these accounts should agree in how they construe the function of morality in human societies.

If this is correct, there will be some limit to how weird and wonderful alien moralities could be in terms of the obligations they impose, the values they honour,

[7] Beckner (1959); Burge (1989); Nanay (2010); Neander (1991b).
[8] See Godfrey-Smith (1993, 1994), Griffiths (1993), Millikan (1984b, 1989), Neander (1991a, 1991b), Wright (1973). Cf. Buller (1998).
[9] *E.g.*, Boorse (1976, 2002), Cummins (1975), Lewens (2004).

and so forth. In particular, this line of reasoning might well vindicate the suggestion that moralities must prescribe some degree of benevolence and/or impartiality. In that respect, the phyletic contingency of our morality would be far from total. The scope of evolutionary debunking arguments would be similarly limited.

Bibliography

Abarbanell, Linda & Hauser, Marc D. (2010) Mayan morality: a study in permissible harms. *Cognition* 115, 207-224.

Aharoni, Eyal, Walter Sinnott-Armstrong, & Kent A. Kiehl (2012) Can psychopathic offenders discern moral wrongs? A new look at the moral/conventional distinction. *Journal of Abnormal Psychology* 121, 484-497.

Alexander, Joshua & Weinberg, Jonathan M. (2007) Analytic epistemology and experimental philosophy. *Philosophy Compass* 2, 56-80.

Alexander Richard D. (1987) *The biology of moral systems*. Hawthorne, NY: Aldine De Gruyter.

Anderson, Steven W., Antoine Bechara, Hanna Damasio, Daniel Tranel, & Antonio R. Damasio (1999) Impairment of social and moral behavior related to early damage

in human prefrontal cortex. *Nature Neuroscience* 2, 1032-1037.

Anonymous (1871) Darwin on *The Descent of Man. The Edinburgh Review* 134, 195-235.

Appiah, Kwame A. (2008) *Experiments in ethics*. Cambridge, MA: Harvard University Press.

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

Asch, Solomon (1952) Group forces in the modification and distortion of judgments. In Guy E. Swanson, Theodore M. Newcomb, & Eugene L. Hartley, eds. (1952) *Readings in social psychology, 2nd ed.*, 2–11. New York: NY Holt.

Audi, Robert (1997a) Ethical naturalism and the explanatory power of moral concepts. In his *Moral knowledge and ethical character*, 112-128. Oxford: Oxford University Press.

(1997b) The place of testimony in the fabric of knowledge and justification. *American Philosophical Quarterly* 34, 405–422.

Aune, Bruce (1991) *Knowledge of the external world*. London: Routledge.

Axelrod, Robert (1980a) Effective choice in the Prisoner's Dilemma. *Journal of Conflict Resolution* 24, 3-25.

(1980b) More effective choice in the Prisoner's Dilemma. *Journal of Conflict Resolution* 24, 379-403.

Axelrod, Robert & Hamilton, William D. (1981) The evolution of cooperation. *Science* 211, 1390-1396.

Ayer, A. J. (1956) *The problem of knowledge*. London: Macmillan.

Baker Mark C. (2001) *The atoms of language: the mind's hidden rules of grammar*. Oxford: Oxford University Press.

Barclay, Pat (2006) Reputational benefits for altruistic punishment. *Evolution and*

Human Behavior 27, 325-344.

Barnes, E. C. (2000) Ockham's Razor and the Anti-Superfluity Principle. *Erkenntnis* 53, 353-374.

Barrett, Justin L. (2000) Exploring the natural foundations of religion. *Trends in Cognitive Sciences* 4, 29-34.

Barrett, Justin L. & Nyhof, Melanie A. (2001) Spreading non-natural concepts: the role of intuitive conceptual structures in memory and transmission of cultural materials. *Journal of Cognition and Culture* 1, 69-100.

Barrett, Lisa Feldman (2006) Are emotions natural kinds? *Perspectives on Psychological Science* 1, 28-58.

Bateson, Patrick (1991) Are there principles of behavioural development? In Patrick Bateson, ed. (1991) *The development and integration of behaviour: essays in honour of Robert Hinde*, 19-39. Cambridge: Cambridge University Press.

Beatty, John (1995) The evolutionary contingency thesis. In Gereon Walters & James Lennox, eds. (1995) *Concepts, theories, and rationality in the biological sciences*, 45-81. Pittsburgh, PA: University of Pittsburgh Press.

Becker, Kelly (2007) *Epistemology modalized*. London: Routledge.

(2012) Methods and how to individuate them. In Kelly Becker & Tim Black, eds. (2012) *The sensitivity principle in epistemology*, 81-97. Cambridge: Cambridge University Press.

Beckner, Morton (1959) *The biological way of thought*. New York, NY: Columbia University Press.

Bergmann, Michael (1997) Internalism, externalism, and the no-defeater condition. *Synthese* 110, 399-417.

(2006) *Justification without awareness: a defense of epistemic externalism*. Oxford: Oxford University Press.

Berker, Selim (2009) The normative insignificance of neuroscience. *Philosophy and Public Affairs* 37, 293-329.

Bittles, Alan H. (1990) Consanguineous marriage: current global incidence and its relevance to demographic research. *Population Research Center*, Report number 90-186. Ann Arbor, MI: University of Michigan.

Blackburn, Simon (1984) *Spreading the word: groundings in the philosophy of language.* Oxford: Oxford University Press.

(1993) *Essays in Quasi-Realism.* Oxford: Oxford University Press.

(1996) Securing the nots: moral epistemology for the Quasi-Realist. In Walter Sinnott-Armstrong & Mark Timmons, eds. (1996) *Moral knowledge? New readings in moral epistemology*, 82-100. Oxford: Oxford University Press.

(1998) *Ruling passions: a theory of practical reasoning.* Oxford: Oxford University Press.

(2010) The steps from doing to saying. *Proceedings of the Aristotelian Society* 110, 1-13.

Blackmore, Susan (1999) *The meme machine.* Oxford: Oxford University Press.

Blair, James, Derek Mitchell, & Karina Blair (2005) *The psychopath: emotion and the brain.* Oxford: Blackwell.

Blair, Robert James Richard (1995) A cognitive developmental approach to morality: investigating the psychopath. *Cognition* 57, 1-29.

(1997) Moral reasoning in the child with psychopathic tendencies. *Personality and Individual Differences* 11, 620-628.

(1999) Responsiveness to distress cues in children with psychopathic tendencies. *Personality and Individual Differences* 27, 135 – 145.

Blair, Robert James Richard, Lawrence Jones, Fiona Clark, & Margaret Smith (1997). The psychopathic individual: A lack of responsiveness to distress cues.

Psychophysiology 34, 192-198.

Boehm, Christopher (1999) *Hierarchy in the forest: the evolution of egalitarian behavior*. Cambridge, MA: Harvard University Press.

(2000) Conflict and the evolution of social control. *Journal of Consciousness Studies* 7, 79-101.

Boghossian, Paul (1996) Analyticity reconsidered. *Noûs* 30, 360-391.

(2003) Epistemic analyticity: a defense. *Grazer Philosophische Studien* 56, 15-35.

Boinski, Sue (1994) Affiliation patterns among male Costa Rican squirrel monkeys. *Behaviour* 130, 191-209.

Bond, Rob & Smith, Peter B. (1996) Culture and conformity: a meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin* 119, 111-137.

BonJour, Laurence (1980) Externalist theories of knowledge. *Midwest Studies in Philosophy* 5, 53-74.

Boorse, Cristopher (1976) Wright on functions. *The Philosophical Review* 85, 70-86.

(2002) A rebuttal on functions. In André Ariew, Robert Cummins, & Mark Perlman, eds. (2002) *Functions: new essays in the philosophy of psychology and biology*, 7-32. Oxford: Oxford University Press.

Bostrom, Nick (2002). *Anthropic bias: observation selection effects in science and philosophy*. London: Routledge.

Boulter, Stephen J. (2007) The 'evolutionary argument' and the metaphilosophy of commonsense. *Biology and Philosophy* 22, 369-382.

Bourget, David & Chalmers, David J. (2010) *The PhilPapers surveys: results*. Last accessed 24.01.2014: <http://philpapers.org/surveys/results.pl>

Bowler, Peter J. (1989) *Evolution: the history of an idea*. London: University of California Press.

Bowles, Samuel (2006) Group competition, reproductive levelling, and the evolution of

human altruism. *Science* 314, 1569-1572.

(2009) Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science* 324, 1293-1298.

Bowles, Samuel & Gintis, Herbert (2011) *A cooperative species: human reciprocity and its evolution.* Princeton, NJ: Princeton University Press.

Boyd, Kenneth & Nagel, Jennifer (forthcoming). The reliability of epistemic intuitions. In Edouard Machery, ed. (forthcoming) *Current controversies in experimental philosophy.* London: Routledge.

Boyd, Richard N. (1988) How to be a moral realist. In Geoffrey Sayre-McCord, ed. (1988) *Essays on moral realism*, 307-356. Ithaca, NY: Cornell University Press.

Boyd, Robert & Richerson, Peter J. (1985) *Culture and the evolutionary process.* Chicago, IL: University of Chicago Press.

(1992) Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology* 13, 171-195.

Boyer, Pascal (2001) *Religion explained: the human instincts that fashion gods, spirits and ancestors.* London: Vintage Books.

Boyer, Pascal & Ramble, Charles (2001) Cognitive templates for religious concepts: cross-cultural evidence for recall of counterintuitive representations. *Cognitive Science* 25, 535-564.

Brandon, Robert N. (1989) *Adaptation and environment.* Princeton, NJ: Princeton University Press.

Brandt, Richard B. (1954) *Hopi ethics: a theoretical analysis.* Chicago, IL: University of Chicago Press.

(1959) *Ethical theory: the problems of normative and critical ethics.* New York, NY: Prentice Hall.

Briggs, Rachael (2009) Distorted reflection. *The Philosophical Review* 118, 59-85.

Brink, David (1989) *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.

Brogaard, Berit (2008) Moral contextualism and moral relativism. *Philosophical Quarterly* 58, 385-409.

(2012) Moral relativism and moral expressivism. *The Southern Journal of Philosophy* 50, 538-556.

Brooks, William Keith (1899) *Foundations of zoology*. London: Macmillan.

Broome, John (2005) Should we value population? *Journal of Political Philosophy* 13, 399-413.

Brosnan, Kevin (2011) Do the evolutionary origins of morality undermine moral knowledge? *Biology and Philosophy* 26, 51-64.

Brosnan, Sarah F., Hillary C. Schiff, & Frans B. M. de Waal (2005) Tolerance for inequity may increase with social closeness in chimpanzees. *Proceedings of the Royal Society B* 272, 253–258.

Brown, Donald (1991) *Human universals*. New York, NY: McGraw-Hill.

Buckwalter, Wesley, & Stich, Stephen (forthcoming) Gender and philosophical intuition. In Joshua Knobe & Shaun Nichols, eds. (forthcoming) *Experimental philosophy, Vol* 2. Oxford: Oxford University Press.

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

Burge, Tyler (1989) Individuation and causation in psychology. *Pacific Philosophical Quarterly* 4, 303-322.

(1993) Content preservation. *Philosophical Review* 102, 457–488.

Buss, David (2000) *The dangerous passion: why jealousy is as necessary as love and sex*. New York, NY: The Free Press.

Calcott, Brett (2013) Why how and why aren't enough: more problems with Mayr's

proximate-ultimate distinction. *Biology and Philosophy*, online preprint.

Camerer, Colin F. (2003) *Behavioral game theory: experiments in strategic interaction.* Princeton, NJ: Princeton University Press.

Cappelen, Herman & Hawthorne, John (2009) *Relativism and monadic truth.* Oxford: Oxford University Press.

Carlsmith, Kevin M., John M. Darley, & Paul H. Robinson (2002) Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology* 83, 284-299.

Carroll, Noël (1996) Moderate moralism. *British Journal of Aesthetics* 36, 223-238.

Carruthers, Peter (1992) *Human knowledge and human nature: a new introduction to an ancient debate.* Oxford: Oxford University Press.

Carruthers, Peter & James, Scott M. (2008) Human evolution and the possibility of moral realism. *Philosophy & Phenomenological Research* 77, 237-244.

Carter, Brandon (1974) Large number coincidences and the anthropic principle in cosmology. In M. S. Longair, ed. (1974) *Confrontation of cosmological theories with observational data*, 291-298. Dordrecht: Kluwer.

Casullo, Albert (2003) *A priori justification.* Oxford: Oxford University Press.

Cavalli-Sforza, Luigi & Feldman, Marcus W. (1981) *Cultural transmission and evolution: a quantitative approach.* Princeton, NJ: Princeton University Press.

Chomsky, Noam (1981) *Lectures on government and binding: the Pisa lectures.* Dordrecht: Foris.

(1995) *The minimalist program.* Cambridge: Cambridge University Press.

Christensen, David (2007) Epistemology of disagreement: the good news. *Philosophical Review* 116, 187-217.

(2010) Disagreement, question-begging, and epistemic self-criticism. *Philosophers' Imprint* 11, 1-22.

Chudek, Maciek, Wanying Zhao, & Joseph Henrich (2013) Culture-gene coevolution, large-scale cooperation, and the shaping of human social psychology. In Kim Sterelny, Richard Joyce, Brett Calcott, & Ben Fraser, eds. (2013) *Cooperation and its evolution*, 425-458. Cambridge, MA: MIT Press.

Ciaramelli, Elisa, Michela Muccioli, Elisabetta Ladavas, & Giuseppe di Pellegrino (2007) Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience* 2, 84–92.

Clark, Andy (2005) Intrinsic content, active memory, and the extended mind. *Analysis* 65, 1-11.

Clarke, Samuel (1738/2003) A discourse concerning the unchangeable obligations of natural religion. In J. B. Schneewind, ed. (2003) *Moral philosophy from Montaigne to Kant*, 295-311. Cambridge: Cambridge University Press.

Coady, C. A. J. (1992) *Testimony: a philosophical study*. Oxford: Clarendon Press.

Cobbe, Frances (1871) Darwinism in morals. *The Theological Review* 8, 167-192.

Cohen, Gerald A. (2000) *If you're an egalitarian, how come you're so rich?* Cambridge, MA: Harvard University Press.

Cohen, Stewart (1984) Justification and truth. *Philosophical Studies* 46, 279–295.

(2010) Bootstrapping, defeasible reasoning, and *a priori* justification. *Philosophical Perspectives* 24, 141-159.

Collins, Robin (2009) The Teleological Argument: an exploration of the fine-tuning of the universe. In William Lane Craig & J. P. Moreland, eds. (2009) *The Blackwell Companion to Natural Theology*, 202-281. Oxford: Blackwell.

Colyvan, Mark (1998) Can the Eleatic Principle be justified? *Canadian Journal of Philosophy* 28, 313-335.

Colyvan, Mark, Jay L. Garfield, & Graham Priest (2005) Problems with the Argument from Fine Tuning. *Synthese* 145, 325-338.

Comesaña, Juan (2002) The diagonal and the demon. *Philosophical Studies* 110, 249–266.

Cooper, David E. (1978) Moral relativism. *Midwest Studies in Philosophy* 3, 97-108.

Copp, David (1990) Explanation and justification in ethics. *Ethics* 100, 237-258.

(1995) *Morality, normativity, and society*. Oxford: Oxford University Press.

(2007) *Morality in a natural world: selected essays in metaethics*. Cambridge: Cambridge University Press.

(2008) Darwinian skepticism about moral realism. *Philosophical Issues* 18, 186-206.

(2009) Toward a pluralist and teleological theory of normativity. *Philosophical Issues* 19, 21-37.

(2012) Normativity and reasons: five arguments from Parfit against normative naturalism. In Susanna Nuccetelli & Gary Seay, eds. (2012) *Ethical naturalism: current debates*, 24-57. Cambridge: Cambridge University Press.

Corfield, David (2004) Mathematical kinds, or being kind to mathematics. *Philosophica* 74, 37–62.

Cosmides, Leda, H. Clark Barrett, & John Tooby (2010) Adaptive specializations, social exchange, and the evolution of human intelligence. *Proceedings of the National Academy of Sciences* 107(suppl. 2), 9007-9014.

Cosmides, Leda, & Tooby, John (1992) Cognitive adaptations for social exchange. In Jerome H. Barkow, Leda Cosmides, & John Tooby, eds. (1992) *The adapted mind: evolutionary psychology and the generation of culture*, 163-228. New York: Oxford University Press.

(2008) Can a general deontic logic capture the facts of human moral reasoning? How the mind interprets social exchange rules and detects cheaters. In Walter Sinnott-Armstrong, ed. (2008) *Moral psychology, vol. 1: the evolution of morality*, 53-120. Cambridge, MA: MIT Press.

Cowie, Fiona (1999) *What's within? Nativism reconsidered*. Oxford: Oxford University Press.

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

Cudworth, Ralph (1731/1996) *A treatise concerning eternal and immutable morality*. Cambridge: Cambridge University Press.

Cummins, Robert (1975) Functional analysis. *Journal of Philosophy* 72, 741-764.

Cuneo, Terence (2006) Moral facts as configuring causes. *Pacific Philosophical Quarterly* 87, 141-162.

Curry, Andrew (2013) The milk revolution. *Nature* 500, 20-22.

Dancy, Jonathan (1985) *An introduction to contemporary epistemology*. Oxford: Blackwell.

(2004) *Ethics without principles*. Oxford: Oxford University Press.

Darwin, Charles (1872/2009) *The expression of the emotions in man and animals*. London: Penguin.

(1879/2004) *The descent of man, and selection in relation to sex, 2nd ed.* London: Penguin

Davidson, Donald (1969) On saying that. *Synthese* 19, 130-146.

(1973) Radical interpretation. *Dialectica* 27, 313-328.

(1974) On the very idea of a conceptual scheme. *Proceedings and Addresses of the American Philosophical Association* 47, 5-20.

(1995) The objectivity of values. In his (2004) *Problems of rationality*, 39-57. Oxford: Oxford University Press.

Davies, Nicholas B., John R. Krebs, & Stuart A. West (2012) *An introduction to behavioural ecology, 4th ed.* Chichester: Wiley-Blackwell.

Davies, Paul (1992) *The mind of God: the scientific basis for a rational world*. New York: Touchstone.

Davis, Philip J. (1981) Are there coincidences in mathematics? *The American Mathematical Monthly* 88, 311-320.

Dawkins, Richard (1976) *The selfish gene*. Oxford: Oxford University Press.

(2006) *The God delusion*. London: Bantam Press.

De Dreu, Carsten K. W., Lindred L. Greer, Gerben A. Van Kleef, Shaul Shalvi, & Michel J. J. Handgraaf (2011) Oxytocin promotes human ethnocentrism. *Proceedings of the National Academy of Sciences* 108, 1262-1266.

de Finetti, Bruno (1974) *Theory of probability, vol. 1*, transl. Einaudi. London: Wiley.

de Lazari-Radek, Katarzyna and Singer, Peter (2012) The objectivity of ethics and the unity of practical reason. *Ethics* 123, 9-31.

de Waal, Frans B. M. (1982) *Chimpanzee politics*. London: Jonathan Cape.

(1989) Food sharing and reciprocal obligations among chimpanzees. *Journal of Human Evolution* 18, 433-459.

(1991) The chimpanzee's sense of social regularity and its relation to the human sense of justice. *American Behavioral Scientist* 34, 335-349.

(1996) *Good natured: the origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press.

(2001) *The ape and the sushi master: cultural reflections of a primatologist*. New York, NY: Basic Books.

(2006). *Primates and philosophers: how morality evolved*. Princeton, NJ: Princeton University Press.

de Waal, Frans B. M., & Aureli, Filippo (1996). Consolation, reconciliation, and a possible cognitive difference between macaques and chimpanzees. In Anne E. Russon, Kim A. Bard, & Sue Taylor Parker, eds. (1996) *Reaching in thought: the minds of the Great Apes*, 80-110. Cambridge: Cambridge University Press.

de Waal, Frans B. M. & Flack, Jessica (2000) 'Any animal whatever' Darwinian

building blocks of morality in monkeys and apes. *Journal of Consciousness Studies* 7, 1–29.

de Waal, Frans B. M & van Roosmalen, Angeline (1979). Reconciliation and consolation among chimpanzees. *Behavioral Ecology & Sociobiology* 5, 55-66.

Dean, L. G., R. L. Kendal, S. J. Shapiro, B. Thierry, & K. N. Laland (2012) Identification of the social and cognitive processes underlying human cumulative culture. *Science* 335, 1114-1118.

Dennett, Daniel (1995) *Darwin's dangerous idea: evolution and the meanings of life*. New York: Touchstone.

DeRose, Keith (1995) Solving the skeptical problem. *Philosophical Review* 104, 1-52.

Descartes, Rene (1641/1996) *Meditations on first philosophy, with selections from the objections and replies*, transl. Cottingham. Cambridge: Cambridge University Press.

Doris, John M. & Plakias, Alexandra (2008) How to argue about disagreement: evaluative diversity and moral realism. In Walter Sinnott-Armstrong, ed. (2008) *Moral psychology, vol. 2: the cognitive science of morality*, 303-331. Cambridge, MA: MIT Press.

Doris, John M. & Stich, Stephen (2005) As a matter of fact: empirical perspectives on ethics. In Frank Jackson & Michael Smith, eds. (2005) *The Oxford handbook of contemporary analytic philosophy*, 114-152. Oxford: Oxford University Press.

Draghi-Lorenz, Riccardo, Vasudevi Reddy, & Alan Costall (2001) Rethinking the development of 'nonbasic' emotions: a critical review of existing theories. *Developmental Review* 21, 263–304.

Dreber, Anna, David G. Rand, Drew Fudenburg, & Martin A. Nowak (2008) Winners don't punish. *Nature* 452, 348-351.

Dreier, James (1990) Internalism and speaker relativism. *Ethics* 101, 6-26.

Dretske, Fred (1970) Epistemic operators. *Journal of Philosophy* 67, 1007-1023.

(2005) Is knowledge closed under known entailment? The case against closure. In Matthias Steup & Ernest Sosa, eds. (2005) *Contemoprary Debates in Epistemology*, 13-26. Oxford: Blackwell.

Driver, Julia (2006) Autonomy and the asymmetry for moral expertise. *Philosophical Studies* 128, 619 – 644.

Durham, William H. (1991) *Coevoluton: genes, cultures, and human diversity*. Stanford, CA: Stanford University Press.

Dworkin, Ronald (1996) Truth and objectivity: you'd better believe it. *Philosophy and Public Affairs* 25, 87-139.

Dwyer, Susan (1999) Moral competence. In Kumiko Murasugi & Robert Stainton, eds. (1999) *Philosophy and linguistics*, 169-190. Boulder, CO: Westview Press.

(2006) How good is the linguistic analogy? In Peter Carruthers, Stephen Laurence, & Stephen Stich, eds. (2006) *The innate mind, vol. 2: culture and cognition*, 237-256. Oxford: Oxford University Press.

Dwyer, Susan, Bryce Huebner, & Marc D. Hauser (2010) The linguistic analogy: motivations, results, and speculations. *Topics in Cognitive Science* 2, 486-510.

Eagle, Anthony (2012) Some basic principles about chance. In Edward N. Zalta, ed. (2013) *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition). Last accessed 23.10.2013: <http://plato.stanford.edu/archives/spr2013/entries/chance-randomness/ basic-chance.html>

Egas, Martijn & Riedl, Arno (2008) The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B* 275, 871-878.

Eidelman, Scott, Christian S. Crandall, Jeffrey A. Goodman, & John C. Blanchar (2012) Low-effort thought promotes conservatism. *Personality and Social Psychology Bulletin* 4, 316-323.

Ekman, Paul (1972) Universals and cultural differences in facial expressions of

emotions. *Nebraska Symposium on Motivation* 19, 207-283.

(1992) An argument for basic emotions. *Cognition and Emotion* 6, 169-200.

Ekman, Paul & Friesen, Wallace V. (1971) Constants across culture in the face and emotion. *Journal of Personality and Social Psychology* 17, 124-129.

Elga, Adam (2007) Reflection and disagreement. *Nous* 41, 478-502.

(ms.) Lucky to be rational. Last accessed 22.07.2013: <http://www.princeton.edu/~adame/papers/bellingham-lucky.pdf>.

Ember, Carol R. (1978) Myths about hunter gatherers. *Ethnology* 17, 439-448.

Enoch, David (2009) How is moral disagreement a problem for realism? *Journal of Ethics* 13, 15-50.

(2010a) The epistemological challenge to metanormative realism: how best to understand it, and how to cope with it. *Philosophical Studies* 148, 413-438.

(2010b) Not just a truthometer: taking oneself seriously (but not too seriously) in cases of peer disagreement. *Mind* 119, 953-997.

(2011) *Taking morality seriously*. Oxford: Oxford University Press.

Ertan, Arhan, Talbot Page, & Louis Putterman (2009) Who to punish? Individual decisions and majority rule in the solution of free rider problems. *European Economic Review* 3, 495-511.

Fales, Evan (2002) Darwin's doubt, Calvin's Calgary. In James Beilby, ed. (2002) *Naturalism defeated? Essays on Plantinga's evolutionary argument against naturalism*, 43-58. Ithaca, NY: Cornell University Press.

Fehr, Ernst & Gächter, Simon (2002). Altruistic punishment in humans. *Nature* 415, 137-140.

Fehr, Ernst & Henrich, Joseph (2003) Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. In Peter Hammerstein, ed. (2003) *Genetic and cultural evolution of cooperation*, 55-82. New York, NY: MIT Press.

Feldman, Marcus W. & Cavalli-Sforza, Luigi (1989) On the theory of evolution under genetic and cultural transmission with application to the lactose absorption problem. In Marcus W. Feldman, ed. (1989) *Mathematical evolutionary theory*, 145-173. Princeton, NJ: Princeton University Press.

Feltz, Edward & Cokely, Edward T. (2012) The philosophical personality argument. *Philosophical Studies* 161, 227-246.

Field, Hartry (1998) Mathematical objectivity and mathematical objects. In his (2001) *Truth and the absence of fact*, 315-331. Oxford: Oxford University Press.

Fiske, Alan P. (1992) The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological Review* 99, 689-723.

FitzPatrick, William (2008) Morality and evolutionary biology. In Edward N. Zalta, ed. (2008) *The Stanford Encyclopedia of Philosophy* (Winter 2008 Edition). Last accessed 12.11.2013: <http://plato.stanford.edu/archives/win2008/entries/morality-biology/>

Fodor, Jerry (1981) Three cheers for propositional attitudes. In Jerry Fodor (1981) *RePresentations: philosophical essays on the foundations of cognitive science*, 100-123.Cambridge, MA: MIT Press.

(2002) Is science biologically possible? In James Beilby, ed. (2002) *Naturalism defeated? Essays on Plantinga's evolutionary argument against naturalism*, 30-42. Ithaca, NY: Cornell University Press.

Foley, Richard (2001) *Intellectual trust in oneself and others*. Cambridge: Cambridge University Press.

Foot, Philippa (1958) Moral arguments. *Mind* 67, 502-513.

(1967) The problem of abortion and the Doctrine of Double Effect. *Oxford Review* 5, 5–15.

Frank, Robert (2003) Repression of competition and the evolution of cooperation.

Evolution 57, 693-705.

Fraser, Ben & Hauser, Marc D. (2010) The argument from disagreement and the role of cross-cultural empirical data. *Mind and Language* 25, 451-560.

Fricker, Elizabeth (1987) The epistemology of testimony. *Proceedings of the Aristotelian Society, Suppl. Volume* 61, 57-83.

(1994) Against gullibility. In Bimal K. Matilal & Arindam Chakrabati, eds. (1994) *Knowing from words: Western and Indian philosophical analysis of understanding and testimony*, 125-161. Dordrecht: Kluwer.

(1995) Critical notice: Telling and trusting: reductionism and anti-reductionism in the epistemology of testimony. *Mind* 104, 393–411.

Gangestad, Steven W. (2006) Evidence for adaptations for female extra-pair mating in humans: thoughts on current status and future directions. In Steven M. Platek & Todd K. Shackelford, eds. (2006) *Female infidelity and paternal uncertainty: evolutionary perspectives on male anti-cuckoldry tactics*, 37-57. Cambridge: Cambridge University Press.

Garfinkel, Alan (1981) *The forms of explanation.* New Haven, Connecticut: Yale University Press.

Gaut, Berys (2007) *Art, emotion, and ethics.* Oxford: Oxford University Press.

Geraci, Alessandra & Surian, Luca (2011) The developmental roots of fairness: infants' reactions to equal and unequal distributions of resources. *Developmental Science* 14, 1012-1020.

Gibbard, Alan (1990) *Wise choices, apt feelings: a theory of normative judgment.* Cambridge, MA: Harvard University Press.

(2003) *Thinking how to live.* Cambridge, MA: Harvard University Press.

(2011) How much realism? Evolved thinkers and normative concepts. In Russ Shafer-Landau, ed. (2011) *Oxford studies in metaethics, vol. 6*, 33-51. Oxford: Oxford

University Press.

Gil-White, Francisco J. (1999) How thick is blood? The plot thickens . . . : If ethnic actors are primordialists, what remains of the circumstantialists/primordialists controversy?' *Ethnic and Racial Studies* 22, 789–820.

(2001) Are ethnic groups biological 'species' to the human brain?' *Current Anthropology* 42, 515–554.

Gilbert, Matthew (2004) Police seeing double in rape case involving identical twins. *CNN.com*. Last accessed 22.07.2013: <http://edition.cnn.com/2004/LAW/06/07/twins/>

Gintis, Herbert (2013) Territoriality and loss aversion: the evolutionary roots of property rights. In Kim Sterelny, Richard Joyce, Brett Calcott, & Ben Fraser, eds. (2013) *Cooperation and its evolution*, 117-130. Cambridge, MA: MIT Press.

Gintis, Herbert, Eric Alden Smith, & Samuel Bowles (2001) Costly signalling and cooperation. *Journal of Theoretical Biology* 213, 103-119.

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

(1994) A modern history theory of functions. *Nous* 28, 344-362.

Goldman, Alan H. (1988) *Empirical knowledge*. Berkeley, CA: University of California Press.

Goldman, Alvin I. (1979) What is justified belief? In George Pappas, ed. (1979) *Justification and knowledge*, 1-23. Dordrecht: D. Reidel.

(1986) *Epistemology and cognition*, Cambridge, MA: Harvard University Press.

Goodall, Jane (1982) Order without law. *Journal of Social and Biological Structures* 5, 349–52.

(1986) *The chimpanzees of Gombe: patterns of behavior*. Cambridge, MA: Harvard University Press, Harvard University Press.

(1990) *Through a window: my thirty years with the chimpanzees of Gombe*. Boston: Houghton Mifflin.

Gould, Stephen Jay (1990) *Wonderful life: the Burgess Shale and the nature of history*. London: Vintage.

(1994) Hooking Leviathan by its past. In his (2011) *Dinosaur in a haystack: reflections in natural history*, 359-376. Cambridge, MA: Harvard University Press.

Graham, Joseph, Jonathan Haidt, & Brian A. Nosek (2009) Liberals and conservatives rely on different sets of moral foundations. *Personality Processes and Individual Differences* 96, 1029-1046.

Greene, Joshua D. (2008) The secret joke of Kant's soul. In Walter Sinnott-Armstrong, ed. (2008) *Moral psychology, vol. 3: the neuroscience of morality*, 35-80. Cambridge, MA: MIT Press.

(2009) Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology* 45, 581-584.

Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John D. Marley, & Jonathan D. Cohen (2001) An fMRI investigation of emotional engagement in moral judgment *Science* 293, 2105-2108.

Griffiths, Paul E. (1993) Functional analysis and proper function. *British Journal for the Philosophy of Science* 44, 409-422.

(1997) *What emotions really are: the problem of psychological categories*. Chicago, IL: Chicago University Press.

(2002) What is innateness? *The Monist* 85, 70-85.

Griffiths, Paul, Edouard Machery, & Stefan Linquist (2009) The vernacular concept of innateness. *Mind and Language* 24, 605-630.

Gross, Neil & Simmons, Solon (2007) The social and political views of American

professors. Unpublished manuscript. Harvard University.

Haidt, Jonathan (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review* 108, 814-834.

(2007) The new synthesis in moral psychology. *Science* 316, 998-1002.

(2012) *The righteous mind: why good people are divided by politics and religion*. London: Penguin.

Haidt, Jonathan & Bjorklund, Fredrik (2008) Social intuitionists answer six questions about morality. In Walter Sinnott-Armstrong, ed. (2008) *Moral psychology, vol. 2: the cognitive science of morality*, 181-218. Cambridge, MA: MIT Press.

Haidt, Jonathan, Fredrik Bjorklund, & Scott Murphy (2000) Moral dumbfounding: when intuitions find no answer. Unpublished manuscript. University of Virginia.

Haidt, Jonathan & Joseph, Craig (2004) Intuitive ethics: how innately prepared intuitions generate culturally variable virtues. *Daedalus* 133, 55-66.

(2007) The moral mind: How 5 sets of innate moral intuitions guide the development of many culture-specific virtues, and perhaps even modules. In Peter Carruthers, Stephen Laurence, & Stephen Stich, eds. (2007) *The innate mind, vol. 3: foundations and the future*, 367-391. Oxford: Oxford University Press.

Haidt, Jonathan & Kesebir, Selin (2010) Morality. In Susan T. Fiske, Daniel T. Gilbert, & Gardner Lindzey, eds. (2010) *Handbook of social psychology, 5th ed.* 797-832. Hoboken, NJ: Wiley.

Haidt, Jonathan, Paul Rozin, Clark McCauley, & Sumio Imada (1997) Body, psyche, and culture: the relationship of disgust to morality. *Psychology and Developing Societies* 9, 107-131.

Hall, Lars, Petter Johansson, & Thomas Strandberg (2012) Lifting the veil of morality: choice-blindness and attitude-reversals on a self-transforming survey. *PLoS One* 7, e45457.

Hamilton, William D. (1964) The genetical evolution of social behaviour I, II. *Journal of Theoretical Biology*7, 1-52.

Hare, Richard M. (1952) *The language of morals*. Oxford: Oxford University Press.

Harker, David (2012) A surprise for Horwich (and some advocates for the fine-tuning argument (which does not include Horwich (as far as I know))). *Philosophical Studies* 161, 247-261.

Harman, Gilbert (1973) *Thought*. Princeton, NJ: Princeton University Press.

(1977) *The nature of morality*. Oxford: Oxford University Press.

(1986) Moral explanations of natural facts: Can moral claims be tested against moral reality? *Southern Journal of Philosophy* 24, suppl., 57-68.

(1996) Moral relativism. In Gilbert Harman & Judith Jarvis Thomson (1996) *Moral relativism and moral objectivity*, 1-64. Oxford: Blackwell.

(1999) Moral philosophy and linguistics. In Gilbert Harman (2000) *Explaining value and other essays in moral philosophy*, 217-226. Oxford: Oxford University Press.

Hart, H. L. A. & Honoré, Anthony (1959) *Causation in the law*. Oxford: Oxford University Press.

Hauser, Marc, Fiery Cushman, Liane Young, R. Kang-Xing Jin, & John Mikhail (2007) A dissociation between moral judgments and justifications. *Mind and Language* 22, 1–21.

Hawking, Stephen (1998) *A brief history of time: from the Big Bang to black holes, 10th ed.* New York: Bantam Books.

Hawthorne, John (2002) Deeply contingent *a priori* knowledge. *Philosophy and Phenomenological Research* 65, 247-269.

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

(2005) The case for closure. In Matthias Steup & Ernest Sosa, eds. (2005) *Contemoprary debates in epistemology*, 26-43. Oxford: Blackwell.

Hawthorne, John & Srinivasan, Amia (2013) Disagreement without transparency: some bleak thoughts. In David Christensen & Jennifer Lackey, eds. (2013) *The epistemology of disagreement: new essays*, 9-30. Oxford: Oxford University Press.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Michael Alvard, Abigail Barr, Natalie Smith Henrich, Kim Hill, Francisco Gil-White, Michael Gurven, Frank W. Marlowe, John Q. Patton, & David Tracer (2005) 'Economic man' in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behavioral and Brain Science* 28, 795–855.

Henrich, Joseph, Steven J. Heine, & Ara Norenzayan (2010) The weirdest people in the world? *Behavioral and Brain Sciences* 33, 61-83.

Henrich, Natalie & Henrich, Joseph (2007) *Why humans cooperate: a cultural and evolutionary explanation.* Oxford: Oxford University Press.

Henrich, Joseph & Silk, Joan B. (2013) Interpretative problems with chimpanzee ultimatum game. *Proceedings of the National Academy of Sciences* 110, E3049.

Herrmann, Benedikt, Christian Thöni, & Simon Gächter (2008) Anti-social punishment across societies. *Science* 319, 1362-1367.

Heyes, Cecelia (2012) Simple minds: a qualified defence of associative learning. *Philosophical Transactions of the Royal Society B* 367, 2695-2703.

Hills, Alison (2010) *The beloved self: morality and the challenge from egoism.* Oxford: Oxford University Press.

Holder, Rodney D. (2004) *God, the Multiverse, and everything: modern cosmology and the Argument from Design.* Aldershot: Ashgate.

Hopkins, Robert (2007) What is wrong with moral testimony? *Philosophy and Phenomenological Research* 74, 611-634.

Horberg, E. J., Christopher Oveis, & Dacher Keltner (2011) Emotions as moral amplifiers: an appraisal tendency approach to the influences of distinct emotions upon moral judgment. *Emotion Review* 3, 237-244.

Horberg, E. J., Christopher Oveis, Dacher Keltner, & Adam B. Cohen (2009) Disgust and the moralization of purity. *Journal of Personality and Social Psychology* 97, 963-976.

Horgan, Terry & Timmons, Mark (1991) New wave moral realism meets moral twin earth. *Journal of Philosophical Research* 16, 447-465.

(2000) Copping out on moral twin earth. *Synthese* 124, 139-152.

(2006) Cognitivist expressivism. In Terry Horgan & Mark Timmons, eds. (2006) *Metaethics after Moore*, 255-298. Oxford: Oxford University Press.

Horner, Victoria, J. Devyn Carter, Malini Suchak, & Frans B. M. de Waal (2011) Spontaneous prosocial choice by chimpanzees. *Proceedings of the National Academy of Sciences* 108, 13847–13851.

Horvath, Joachim (2010) How (not) to react to experimental philosophy. *Philosophical Psychology* 23, 447-480.

Horwich, Paul (1982) *Probability and evidence*. Cambridge: Cambridge University Press.

Howson, Colin (2008) De Finetti, Countable Additivity, consistency and coherence. *British Journal for the Philosophy of Science* 59, 1-23.

Hudson, Richard Ellis, Juliann Eve Aukema, Claude Rispe, & Denis Roze (2002) Altruism, cheating, and anticheater adaptations in cellular slime molds. *The American Naturalist* 160, 31-42.

Huebner, Bryce, Susan Dwyer, & Marc Hauser (2009) The role of emotion in moral psychology. *Trends in Cognitive Sciences* 13, 1-6.

Huemer, Michael (2005) *Ethical intuitionism*. Basingstoke: Palgrave Macmillan.

(2008) Revisionary intuitionism. *Social Philosophy and Policy* 25, 368-392.

(2011) Epistemological egoism and agent-centred norms. In Trent Dougherty, ed. (2011) *Evidentialism and its discontents*, 17-32. Oxford: Oxford University Press.

Hume, David (1748/2007) *An enquiry concerning human understanding*. Oxford: Oxford University Press.

(1779/1993) *Dialogues concerning natural religion*. Oxford: Oxford University Press.

Hurley, Susan L. (1985) Objectivity and disagreement. In Ted Honderich, ed. (1985) *Morality and objectivity: a tribute to J. L. Mackie*, 54-97. London: Routledge & Kegan Paul.

(1989) *Natural reasons: personality and polity*. Oxford: Oxford University Press.

Izard, Carroll (1971) *The face of emotion*. East Norwalk, CT: Appleton-Century-Crofts.

Jackson, Frank (1977) *Perception: a representative theory*. Cambridge: Cambridge University Press.

(1998) *From metaphysics to ethics: a defence of conceptual analysis*. Oxford: Oxford University Press.

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

Jensen, Keith, Josep Call, Michael Tomasello (2007) Chimpanzees are rational maximizers in an ultimatum game. *Science* 318, 107-109.

(2013) Chimpanzee responders still behave like rational maximizers. *Proceedings of the National Academy of Sciences* 110, E1837.

Jensen, Keith, Brian Hare, Josep Call, & Michael Tomasello (2006) What's in it for me? Self-regard precludes altruism and spite in chimpanzees. *Proceedings of the Royal Society B*. 273, 1013–1021.

Johnston, Mark (2001) The authority of affect. *Philosophy and Phenomenological Research* 63, 181-214.

Jones, Karen (1999) Second-hand moral knowledge. *Journal of Philosophy* 96, 55-78.

Joyce, Richard (2000) Darwinian ethics and error. *Biology and Philosophy* 15, 713-732.

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

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

(forthcoming) Evolution, truth-tracking, and moral skepticism. In Bastian Reichardt, ed. *Problems of goodness: new essays in metaethics*.

Kahane, Guy (2011) Evolutionary debunking arguments. *Noûs* 45, 103-125.

Kahane, Guy, Kajta Wiech, Nicholas Shackel, Miguel Farrias, Julian Savulescu, & Irene Tracey (2011) The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience* 7, 393-402.

Kahneman, Daniel (2011) *Thinking, fast and slow*. London: Penguin.

Kanner, Leo (1943) Autistic disturbances of affective contact. *Nervous Child* 2, 217-250.

Kaplan, Hillard, Kim Hill, Kristen Hawkes, & Ana Hurtado (1984) Food sharing among Ache hunter-gatherers of eastern Paraguay. *Current Anthropology* 25, 113-115.

Kelly, Daniel (2011) *Yuck! The nature and moral significance of disgust*. Cambridge, MIT: MIT Press.

Kelly, Thomas (2005) The epistemic significance of disagreement. In Tamar Gendler & John Hawthorne, eds. (2005) *Oxford studies in epistemology, vol. 1*, 167-196. Oxford: Oxford University Press.

Kiehl, Kent A. (2008) Without morals: the cognitive neuroscience of criminal psychopaths. In Walter Sinnott-Armstrong, ed. (2008) *Moral psychology, vol. 3: the neuroscience of morality*, 119-150. Cambridge, MA: MIT Press.

Kitcher, Philip (2006) Biology and ethics. In David Copp, ed. (2006) *The Oxford handbook of ethical theory*, 163-185. Oxford: Oxford University Press.

Kiyonari, Toko & Barclay, Pat (2008) Cooperation in social dilemmas: free riding may be thwarted by second-order reward rather than by punishment. *Journal of*

Personality and Social Psychology 95, 826-42.

Knauft, Bruce B. (1991) Violence and sociality in human evolution. *Current Anthropology* 32, 391-428.

Knobe, Josh & Nichols, Shaun (2008) *Experimental philosophy*. Oxford: Oxford University Press.

Koenigs, Michael, Liane Young, Ralph Adolphs, Daniel Tranel, Fiery Cushman, Marc Hauser, & Antonio Damasio (2007) Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature* 446, 908-911.

Kohlberg, Lawrence (1969) Stage and sequence: the cognitive developmental approach to socialization. In David A. Goslin, ed. (1969) *Handbook of socialization theory and research*, 347–380. Chicago, IL: Rand McNally.

Kölbel, Max (2002) *Truth without objectivity*. London: Routledge.

(2004a) Faultless disagreement. *Proceedings of the Aristotelian Society* 104, 53-73.

(2004b) Indexical relativism versus genuine relativism. *International Journal of Philosophical Studies* 12, 297-313.

Kolmogorov, Andrey N. (1950) *Foundations of the theory of probability*, transl. Morrison. New York: Chelsea Publishing Company.

Koperski, Jeffrey (2005) Should we care about fine-tuning? *British Journal of the Philosophy of Science* 56, 303-319.

Kotzen, Matthew (ms) A formal account of epistemic defeat. Last accessed 12.11.2013: < http://matthewkotzen.net/matthewkotzen.net/Research_files/defeatersweb.pdf>

Kramer, Matthew H. (2009) *Moral realism as a moral doctrine*. Oxford: Wiley-Blackwell.

Kripke, Saul (1981) *Naming and necessity*. Oxford: Blackwell.

Kuklinski, James H., Paul J. Quirk, Jennifer Jerit, David Schwieder, & Robert F. Rich (2000) Misinformation and the currency of democratic citizenship. *The Journal of Politics* 62, 790-816.

Kurzban, Robert, Peter DeScioli, & Erin O'Brien (2007) Audience effects on moralistic punishment. *Evolution and Human Behavior* 28, 75-84.

Kymlicka, Will (1989) *Liberalism, community, and culture*. Oxford: Oxford University Press.

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

(2010) A justificationist view of disagreement's epistemic significance. In Alan Millar, Adrian Haddock, & Duncan Pritchard, eds. (2010) *Social epistemology*, 298-325. Oxford: Oxford University Press.

Ladyman, James & Ross, Don (2007) *Every thing must go: metaphysics naturalized*. Oxford: Oxford University Press.

Laland, Kevin N. & Brown, Gillian R. (2011) *Sense and nonsense: evolutionary perspectives on human behaviour, 2nd ed.* Oxford: Oxford University Press.

Laland, Kevin N., John Odling-Smee, William Hoppitt, & Tobias Uller (2012) More on how and why: cause and effect in biology revisited. *Biology and Philosophy*, online preprint.

Laland, Kevin N., Kim Sterelny, John Odling-Smee, William Hoppitt, & Tobias Uller (2011) Cause and effect in biology revisited: is Mayr's proximate-ultimate distinction still useful? *Science* 334, 1512-1516.

Land, Michael F. & Nilsson, Dan-Eric (2012) *Animal eyes, 2nd ed.* Oxford: Oxford University Press.

Lange, Marc (2010) What are mathematical coincidences (and why does it matter)? *Mind* 119, 307-340.

Lehrman, Daniel S. (1953) Critique of Konrad Lorenz's theory of instinctive behavior. *Quarterly Review of Biology* 28, 337-363.

Leigh, Egbert G. (1977) How does selection reconcile individual advantage with the

good of the group? *Proceedings of the National Academy of Sciences* 74, 4542-4546.

Leimar, Olaf & Hammerstein, Peter (2001) Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society B* 268, 745-753.

Leiter, Brian (2001) Moral facts and best explanations. *Social Philosophy and Policy* 18, 79-101.

Lerner, Jennifer S., Julie H. Goldberg, & Philip E. Tetlock (1998) Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin* 24, 563-574.

Leslie, John (1989) *Universes*. London: Routledge.

Levenson, Robert W. (1994) Human emotions: a functional view. In Paul Ekman & R. J. Davidson, eds. (1994) *The nature of emotion: Fundamental questions*, 123-126. Oxford: Oxford University Press.

(2003) Blood, sweat, and fears: the autonomic architecture of emotion. *Annals of the New York Academy of Sciences* 1000, 348-366.

Levitt, Mary J, Ruth A. Weber, Cherie M. Clark, & Patricia McDonnell (1985) Reciprocity of exchange in toddler sharing behavior. *Developmental Psychology* 21, 122-123.

Lewens, Tim (2001) Sex and selection: a reply to Matthen. *British Journal for the Philosophy of Science* 52, 589-598.

(2004) *Organisms and artifacts: design in nature and elsewhere*. Cambridge, MA: MIT Press.

Lewis, David K. (1973) *Counterfactuals*. Oxford: Blackwell.

(1980) A subjectivist's guide to objective chance. In Richard C. Jeffrey, ed. (1980) *Studies in inductive logic and probability, vol. 2*, 263-294. Berkeley, California: University of California Press.

(1986) Causal explanation. In his (1986) *Philosophical papers, vol. 2*, 214–240. Oxford: Oxford University Press.

Lewontin, Richard Charles (1970). The units of selection. *Annual Reviews of Ecology and Systematics* 1, 1–18.

Liao, Matthew S., Alex Wiegmann, Joshua Alexander, & Gerard Vong (2012) Putting the trolley in order: experimental philosophy and the loop case. *Philosophical Psychology* 25, 661-671

Lieberman, Debra (2008) Moral sentiments relating to incest: discerning adaptations from by-products. In Walter Sinnott-Armstrong, ed. (2008) *Moral psychology, vol. 1: the evolution of morality*, 165-190. Cambridge, MA: MIT Press.

Lieberman, Debra, John Tooby, & Leda Cosmides (2003) Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest. *Proceeding of the Royal Society of London* B 270, 819–826.

(2007) The architecture of human kin detection. *Nature* 445, 727-731.

Lillehammer, Hallvard (2007) *Companions in guilt: arguments for ethical objectivity*. London: Palgrave MacMillan.

Lipton, Peter (1991) *Inference to the best explanation*. London: Routledge.

LoBue, Vanessa, Tracy Nishida, Cynthia Chiong, Judy S. DeLoache, & Jonathan Haidt (2009) When getting something good is bad: even three-year-olds react to inequality. *Social Development* 20, 154-170.

Locke, John (1689/1975) *An essay concerning human understanding*. Oxford: Oxford University Press.

Lord, Charles, Lee Ross, & Mark Lepper (1979) Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37, 2098-2109.

Lumsden, Charles J. & Wilson, Edward O. (1981) *Genes, mind, and culture: the*

coevolutionary process. Cambridge, MA: Harvard University Press.

Macdonald, Kai & Macdonal, Tina Marie (2010) The peptide that binds: a systematic review of oxytocin and its prosocial effects in humans. *Harvard Review of Psychiatry* 18, 1-20.

MacFarlane, John (2005) Making sense of relative truth. *Proceedings of the Aristotelian Society* 105, 321-339.

(2007) Relativism and disagreement. *Philosophical Studies* 132, 17-31.

Machery, Edouard & Mallon, Ron (2010) Evolution of morality. In John M. Doris & the Moral Psychology Research Group, eds. (2010) *The moral psychology handbook*, 3-46. Oxford: Oxford University Press.

Machery, Edouard, Ron Mallon, Shaun Nichols, & Stephen Stich (2004) Semantics, cross-cultural style. *Cognition* 92, B1-B12.

Mackie, J. L (1977) *Ethics: inventing right and wrong*. London: Penguin.

Majors, Brad (2003) Moral explanations and the special sciences. *Philosophical Studies* 113, 121-152.

Mallon, Ron & Stich, Stephen (1990) The odd couple: the compatibility of social construction and evolutionary psychology. *Philosophy of Science* 67, 133-154.

Mallon, Ron & Weinberg, Jonathan M. (2006) Innateness as closed process invariance. *Philosophy of Science* 73, 323-344.

Mameli, Matteo & Bateson, Patrick (2011) An evaluation of the concept of innateness. *Philosophical Transactions of the Royal Society B* 366, 436-443.

Manson, Neil A (2000) There is no adequate definition of 'fine-tuned for life'. *Inquiry* 43, 341-351.

(2009) The Fine-Tuning Argument. *Philosophy Compass* 4, 271-286.

Manson, Neil A. & Thrush, Michael J. (2003) Fine-tuning, multiple universes, and the 'this universe' objection. *Pacific Philosophical Quarterly* 84, 67-83.

Mason, Kelby (2010) Debunking arguments and the genealogy of religion and morality. *Philosophy Compass* 5, 770-778.

Martin, Grace B., & Clark, Russell, D. (1982) Distress crying in neonates: species and peer specificity. *Developmental Psychology* 18, 3-9.

Matsumoto, David, Dacher Keltner, Michelle N. Shiota, Maureen O'Sullivan, & Mark Frank (2008) Facial Expressions of Emotions. In Michael Lewis, Jeanette M. Haviland-Jones, & Lisa Feldman Barrett, eds. (2008) *Handbook of the emotions, 3rd ed.*, 211-234. New York, NY: Guilford Press.

Matthen, Mohan (1999) Evolution, Wisconsin style: selection and the explanation of individual traits. *British Journal of the Philosophy of Science* 50, 143-150.

(2003) Is sex really necessary? And other questions for Lewens. *British Journal of the Philosophy of Science* 54, 297-208.

Maynard-Smith, John (1964) Group selection and kin selection. *Nature* 201, 1145-7.

Maynard-Smith, John & Harper, David (2003) *Animal signals*. Oxford: Oxford University Press.

Mayr, Ernst (1961) Cause and effect in biology. *Science* 134, 1501-1506.

McDowell, John (1994) Knowledge by hearsay. In Bimal K. Matilal & Arindam Chakrabati, eds. (1994) *Knowing from words: Western and Indian philosophical analysis of understanding and testimony*, 195-224. Dordrecht: Kluwer.

McGinn, Colin (1997) *Ethics, evil, and fiction*. Oxford: Oxford University Press.

McGrath, Sarah (2004) Moral knowledge by perception. *Philosophical Perspectives* 18, 209-228.

(2009) The puzzle of pure moral deference. *Philosophical Perspectives* 23, 321-344.

McGrew, Timothy, Lydia McGrew, & Eric Vestrup (2001) Probabilities and the Fine-Tuning Argument: a sceptical view. *Mind* 110, 1027-1038.

McGuire, Jonathan, Robyn Langdon, Max Coltheart, & CatrionaMackenzie (2009) A

reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology* 45, 577-580.

McMahan, Jeff (2002) *The ethics of killing: problems at the margins of life*. Oxford: Oxford University Press.

McNaughton, David & Rawling, Piers (2003) Naturalism and normativity. *Aristotelian Society, Supplementary Volume* 77, 23-45.

Medin, Douglas L. & Atran, Scott (2004) The native mind: biological categorization and reasoning in development and across cultures. *Psychological Review* 111, 960-983.

Melis, Alicia P., Felix Warneken, Keith Jensen, Anna-Claire Schneider, Josep Call, & Michael Tomasello (2011). Chimpanzees help conspecifics obtain food and non-food items. *Proceedings of the Royal Society B*. 278, 1405-1413.

Mendez Mario F., Eric Anderson, & Jill Shapira (2005) An investigation of moral judgment in frontotemporal dementia. *Cognitive and Behavioral Neurology* 18, 193–197.

Mikhail, John (2011) *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. Cambridge: Cambridge University Press.

Mikkelson, Jeffrey M. (2004) Dissolving the Water/Wine Paradox. *British Journal of the Philosophy of Science* 55, 137-145.

Milinski, Manfred, Dirk Semman, Theo C. M. Bakker, & Hans-Jurgen Krambeck (2001) Cooperation through indirect reciprocity: image scoring or standing strategy? *Proceedings of the Royal Society B* 268, 2495-2501.

Mill, John Stuart (1859/1991) *On liberty and other essays*. Oxford: Oxford University Press.

Miller, Geoffrey (2000) *The mating mind: how sexual choice shaped the evolution of human nature*. New York, NY: Doubleday.

Millikan, Ruth G. (1984a) Naturalist reflections on knowledge. *Pacific Philosophical Quarterly* 65, 315-344.

(1984b) *Language, thought and other biological categories*. Cambridge, MA: MIT Press.

(1989) In defense of proper functions. *Philosophy of Science* 56, 288-302.

Mitani, John C. (2006) Reciprocal exchange in chimpanzees and other primates. In P. Kappeler & C. van Schaik, eds. (2006) *Cooperation in primates: mechanisms and evolution*, 101-113. Heidelberg: Springer-Verlag.

Mithen, Steven J. (1990) *Thoughtful foragers: a study of prehistoric decision making*. Cambridge: Cambridge University Press.

Moody-Adams, Michele (1997) *Fieldwork in familiar places: morality, culture, and philosophy*. Cambridge, MA: Harvard University Press.

Moran, Richard (2005) Problems of sincerity. *Proceedings of the Aristotelian Society* 105, 341–361.

Muller, Martin N. & Mitani, John C. (2005) Conflict and cooperation in wild chimpanzees. *Advances in the Study of Behavior* 35, 275-331.

Murphy, Gregory L. (2002) *The big book of concepts*. Cambridge, Massachusetts: MIT Press.

Nagel, Thomas (1976) Moral luck. *Proceedings of the Aristotelian Society, Suppl.* 50, 115-135.

(1978) Ethics without biology. In his (1979) *Mortal questions*, 142-146. Cambridge: Cambridge University Press.

(2012) *Mind and cosmos: why the materialist neo-Darwinian conception of nature is almost certainly false*. Oxford: Oxford University Press.

Nanay, Bence (2010) A modal theory of function. *The Journal of Philosophy* 107, 412-431.

Neander, Karen (1988) What does natural selection explain? Correction to Sober. *Philosophy of Science* 55, 422-426.

(1991a) The teleological notion of function. *Australasian Journal of Philosophy* 4, 454-468.

(1991b) Functions as selected effects: the conceptual analyst's defence. *Philosophy of Science* 2, 168-184.

(1995a) Pruning the Tree of Life. *British Journal for the Philosophy of Science* 46, 59-80.

(1995b) Explaining complex adaptations: a reply to Sober's 'Reply to Neander'. *British Journal for the Philosophy of Science* 46, 583-587.

Neta, Ram (2004) Skepticism, abductivism, and the explanatory gap. *Philosophical Perspectives* 14, 296-325.

Newton-Fisher, Nicholas E. (2006). Female coalitions against male aggression in wild chimpanzees of the Budongo forest. *International Journal of Primatology* 27, 1589-1599.

Nichols, Shaun (2004) *Sentimental rules: on the natural foundations of moral judgment.* Oxford: Oxford University Press.

(2005) Innateness and moral psychology. In Peter Carruthers, Stephen Laurence, & Stephe Stich, eds. (2005) *The Innate mind: structure and contents*, 353-370. Oxford: Oxford University Press.

Nickel, Philip (2001) Moral testimony and its authority. *Ethical Theory and Moral Practice* 4, 253-266.

Nisbett, Richard E. (2003) *The geography of thought: how Asians and Westerners think differently – and why.* New York, NY: The Free Press.

Nolan, Daniel (1997) Quantitative parsimony. *British Journal of the Philosophy of Science* 48, 329-343.

(1998) Impossible worlds: a modest approach. *Notre Dame Journal of Formal Logic* 38, 535-573.

Noonan, Harold (2003) *Personal identity, 2nd ed.* London: Routledge.

Nowak, Martin & Sigmund, Karl (1992) Tit for Tat in heterogeneous populations. *Nature* 355, 250-2543.

(1993) A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* 36, 56-68.

(1994) The alternating Prisoner's Dilemma. *Journal of Theoretical Biology* 168, 219-226.

(1998) Evolution of indirect reciprocity by image scoring. *Nature* 393, 573-577.

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

Nucci, Larry P. (2001) *Education in the moral domain*. Cambridge, Cambridge University Press.

Nucci, Larry P. & Turiel, Elliott (1978) Social interactions and the development of social concepts in preschool children. *Child Development* 49, 400-407.

Nyhan, Brendan & Reifler, Jason (2010) When corrections fail: the persistence of political misperceptions. *Political Behavior* 32, 303-330.

Oddie, Graham (2005) *Value, reality, and desire*. Oxford: Oxford University Press.

Olson, Kristina R. & Spelke, Elizabeth S. (2008) Foundations of cooperation in young children. *Cognition* 108, 222-231.

Okasha, Samir (2001) Why won't the group selection controversy go away? *British Journal of the Philosophy of Science* 52, 25-50.

Ostrom, Elinor, James M. Walker, & Roy Gardner (1992) Covenants with and without a sword: self-governance is possible. *American Political Science Review* 86, 404-417.

Owens, David J. (1998) *Causes and coincidences*. Cambridge: Cambridge University Press.

Page, Talbot, Louis Putterman, & Bulent Unel (2005) Voluntary association in public goods experiments: reciprocity, mimicry, and efficiency. *Economic Journal* 115, 1032-1053.

Panchanathan, Karthik & Boyd, Robert (2003) A tale of two defectors: the importance of standing for the evolution of indirect reciprocity. *Journal of Theoretical Biology* 224, 115-126.

(2004) Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* 432, 499-502.

Panksepp, Jaak (1998) *Affective neuroscience: the foundations of human and animal emotions*. Oxford: Oxford University Press.

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

(2011) *On what matters, vol. 2*. Oxford: Oxford University Press.

Parker, Seymour (1976) The precultural basis of the incest taboo: toward a biosocial theory. *American Anthropologist* 78, 285-305.

Pinker, Steven (2011) *The better angels of our nature: why violence has declined*. London: Viking.

Plantinga, Alvin (1993) *Warrant and proper function*. Oxford: Oxford University Press.

Pollock, John L. (1970) The structure of epistemic justification. *American Philosophical Quarterly* 4, 62-78.

(1986) *Contemporary theories of knowledge*. Oxford: Rowman & Littlefield.

Priest, Graham (1997) Editor's introduction. *Notre Dame Journal of Formal Logic* 38 481-487.

Prinz, Jesse J. (2004) *Gut reactions: a perceptual theory of emotion*. Oxford: Oxford University Press.

(2007) *The emotional construction of morals*. Oxford: Oxford University Press.

(2008) Is morality innate? In Walter Sinnott-Armstrong, ed. (2008) *Moral psychology, vol. 1: the evolution of morality*, 367-406. Cambridge, MA: MIT Press.

Prinz, Jesse J. & Nichols, Shaun (2010) Moral emotions. In John M. Doris & the Moral Psychology Research Group, eds. (2010) *The moral psychology handbook*, 111-146. Oxford: Oxford University Press.

Pritchard, Duncan (2005) *Epistemic luck*. Oxford: Oxford University Press.

Prichard, Duncan & Smith, Matthew (2005) The psychology and philosophy of luck. *New Ideas in Psychology* 22, 1-28.

Proctor, Darby, Rebecca A. Williamson, Frans B. M. de Waal, & Sarah Brosnan (2013) Chimpanzees play the ultimatum game. *Proceedings of the National Academy of Sciences* 110, 2070-2075.

Pryor, James (2000) The Skeptic and the Dogmatist. *Noûs* 34, 517–549.

Pust, Joel (2001a) Against explanationist skepticism regarding philosophical intuitions. *Philosophical Studies* 106, 227-258.

(2001b) Natural selection explanation and origins essentialism. *Canadian Journal of Philosophy* 31, 201-220.

(2004) Natural selection and the traits of individual organisms. *Biology and Philosophy* 19, 765-779.

Quine, Willard Van Orman (1948) On what there is. *Review of Metaphysics* 2, 21-36.

(1951) Two dogmas of empiricism. *Philosophical Review* 60, 20-43.

(1969) Natural kinds. In his (1969) *Ontological relativity and other essays*, 114-138. New York, NY: Columbia University Press.

(1974) *The roots of reference*. La Salle, IL: Open Court.

Quinn, Warren S. (1986) Truth and explanation in ethics. *Ethics* 95, 524-544.

Radcliffe-Richards, Janet (2000) *Human nature after Darwin: a philosophical introduction*.

London: Routledge.

Railton, Peter (1986) Moral realism. *Philosophical Review* 95, 163-207.

Ramsey, William (2002) Naturalism defended. In James Beilby, ed. (2002) *Naturalism defeated? Essays on Plantinga's evolutionary argument against naturalism*, 15-29. Ithaca, NY: Cornell University Press.

Reid, Thomas (1764/1997) *An inquiry into the human mind: on the principles of common sense*. Edinburgh: Edinburgh University Press.

Richards, Robert J. (1987) *Darwin and the emergence of evolutionary theories of mind and behaviour*. Chicago, IL: Chicago University Press.

Richerson, Peter J. & Boyd, Robert (1998) The evolution of human ultrasociality. In Iräneus Eibl-Eibefeldt & Frank Kemp Salter, eds. (1998). *Indoctrinability, ideology, and warfare: evolutionary perspectives*, 71-98. Oxford: Berghahn Books.

(2005) *Not by genes alone: how culture transformed human evolution*. Chicago, IL: University of Chicago Press.

Roberts, Gilbert & Sheratt, Thomas N. (1998) Development of cooperative relationships through increasing investment. *Nature* 394, 175-179.

Rockenbach, Bettina & Milinski, Manfred (2006) The efficient interaction of indirect reciprocity and costly punishment. *Nature* 444, 718–723.

Rosch, Eleanor (1973) Natural categories. *Cognitive Psychology* 4, 328-350.

Rosenberg, Alex (2011) *The atheist's guide to reality: enjoying life without illusions*. New York: W. W. Norton & Co.

Ross, W. D. (1930) *The right and the good*. Oxford: Oxford University Press.

Rottman, Joshua & Keleman, Deborah (2012) Aliens behaving badly: children's acquisition of novel purity-based morals. *Cognition* 124, 356-360.

Roush, Sherrilyn (2007) *Tracking truth: knowledge, evidence, and science*. Oxford: Oxford University Press.

Rozin, Paul, Jonathan Haidt, & Clark R. McCauley (2008) Disgust. In Michael Lewis, Jeanette M. Haviland-Jones, & Lisa Feldman Barrett, eds. (2008) *Handbook of the emotions, 3rd ed.*, 757-776. New York, NY: Guilford Press.

Ruse, Michael (1986) *Taking Darwin seriously: a naturalistic approach to philosophy.* Oxford: Blackwell.

(2006) *Darwinism and its discontents.* Cambridge: Cambridge University Press.

Ruse, Michael & Wilson, E. O. (1986) Moral philosophy as applied science. *Philosophy* 61, 173-192.

Russell, Bertrand (1912) *The problem of philosophy.* London: Penguin.

Sagi, Abraham, & Hoffman, Martin L. (1976) Empathic distress in newborns. *Developmental Psychology* 12, 175-176.

Salmon, Nathan (1982) *Reference and essence.* Oxford: Blackwell.

Samuels, Richard (2002) Nativism in cognitive science. *Mind and Language* 17, 233-265.

(2004) Innateness in cognitive science. *Trends in Cognitive Science* 8, 136-141.

Sarkissian, Hagop, John Park, David Tien, Jennifer Cole Wright, & Joshua Knobe (2011) Folk moral relativism. *Mind & Language* 26, 482-505.

Schafer, Karl (2010) Evolution and normative skepticism. *Australasian Journal of Philosophy* 88, 471-488.

(2012) Assessor relativism and the problem of disagreement. *Southern Journal of Philosophy* 50, 602-620.

Schechter, Joshua (2013) Rational self-doubt and the failure of closure. *Philosophical Studies* 163, 428-452.

(ms.) Luck, rationality, and explanation: a reply to Elga's 'Lucky to be Rational.' Last accessed 28.10.2013: <http://www.brown.edu/Departments/Philosophy/onlinepapers/schechter/Luck

RationalityExplanation.pdf>

Schiffer, Stephen (2003) *The things we mean*. Oxford: Oxford University Press.

Schlesinger, George N. (1991) *The sweep of probability*. Notre Dame, Indiana: University of Notre Dame Press.

Schmitt, Frederick (1999) Social epistemology. In John Greco & Ernest Sosa, eds. (1999) *The Blackwell guide to epistemology*, 354-382. Oxford: Basil Blackwell.

(2002) Testimonial justification: the Parity Argument. *Studies in History and Philosophy of Science Part A* 33, 385-406.

Schnall, Simone, Jonathan Haidt, Gerold L. Clore, & Alexander H. Jordan (2008) Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin* 34, 1096-1109.

Schroeder, Mark (2007) *Slaves of the passions*. Oxford: Oxford University Press.

Schwitzgebel, Eric & Cushman, Fiery (2012) Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind and Language* 27, 135-153.

Seidel, Angelika & Prinz, Jesse (2013) Sound morality: irritating and icky noises amplify judgments in divergent moral domains. *Cognition* 127, 1-5.

Shackelford, Todd K. & Goetz, Aaron T. (2007) Adaptations to sperm competition in humans. *Current Directions in Psychological Science* 16, 47-50.

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

(2012) Evolutionary debunking, moral realism, and moral knowledge. *Journal of Ethics and Social Philosophy* 7, 1-37.

Sher, George (2001) But I could be wrong. *Social Philosophy and Policy* 18, 64-78.

Shweder, Richard (1994) 'You're not sick, you're just in love': emotion as an interpretive system. In Paul Ekman & R. J. Davidson, eds. (1994) *The nature of emotion: fundamental questions*, 32-44. Oxford: Oxford University Press.

Sidgwick, Henry (1872) Review: *Darwinism in Morals, and Other Essays* by Frances Power Cobbe. In Bart Schultz, ed. (1999) *The complete works and select correspondence of Henry Sidgwick, 2nd ed.*, 250. Charlottesville, VA: InteLex Corp.

(1905/2000) Further on the criteria of truth and error. In his (2000) *Essays on ethics and method*, 166-170. Oxford: Oxford University Press.

(1906/1981) *The methods of ethics, 7th ed.* Cambridge: Hacking.

Silk, Joan B., Sarah F. Brosnan, Jennifer Vonk, Joseph Henrich, Daniel J. Povinelli, Amanda S. Richardson, Susan P. Lambeth, Jenny Mascaro, & Steven J. Schapiro (2005) Chimpanzees are indifferent to the welfare of unrelated group members. *Nature* 437, 1357–1359.

Simner, Marvin L. (1971) Newborn's responses to the cry of another infant. *Developmental Psychology* 5, 136-150.

Sinclair, Neil (2006) The moral belief problem. *Ratio* 19, 249-260.

Singer, Peter (1976) All animals are equal. In Tom Regan & Peter Singer, eds. (1989) *Animal rights and human obligations, 2nd ed.*, 148-162. New Jersey, NJ: Pearson.

(1981) *The expanding circle: ethics and sociobiology*. Oxford: Clarendon Press.

(2005) Ethics and intuitions. *Journal of Ethics* 9, 331 - 352.

(2006) Review: Richard Joyce, *The evolution of morality*. *Notre Dame Philosophical Reviews*. Last accessed 24.10.2013: <http://ndpr.nd.edu/news/25012/?id=6383>.

Sinhababu, Neil (ms.) The epistemic argument for hedonism. Last accessed 23.10.2013: <http://philpapers.org/archive/SINTEA-3>.

Sinnott-Armstrong, Walter (2002) Moral relativity and intuitionism. *Noûs* 36, 305 - 328.

(2006a) *Moral Skepticisms*. Oxford: Oxford University Press.

(2006b) Moral intuitionism meets empirical psychology. In Terry Horgan & Mark Timmons, eds. (2006) *Metaethics after Moore*, 339-366. Oxford: Oxford University Press.

(2011) Emotion and reliability in moral psychology. *Emotion Review* 3, 288-289.

Skarsaune, Knut Olav (2011) Darwin and moral realism: survival of the iffiest. *Philosophical Studies* 152, 229-243.

Sliwa, Paulina (2012) In defense of moral testimony. *Philosophical Studies* 158, 175-195.

Sloane, Stephanie, Renee Baillargeon, & David Premack (2012) Do infants have a sense of fairness? *Psychological Science* 23, 196-204.

Smetana, Judith G. (1984) Toddlers' social interactions regarding moral and conventional transgressions. *Child Development* 55, 1767-1776.

(1989) Toddlers' social interactions in the context of moral and conventional transgressions in the home. *Developmental Psychology* 25, 499-508.

Smith, Michael (1994) *The moral problem*. Oxford: Blackwell.

Smith, Philip & Silberberg, Alan (2010) Rational maximizing by humans (*Homo sapiens*) in an ultimatum game. *Animal Cognition* 13, 671-677.

Smolin, Lee (1997) *The life of the cosmos*. Oxford: Oxford University Press.

Sober, Elliott (1984) *The nature of selection: evolutionary theory in philosophical focus*. Chicago, IL: University of Chicago Press.

(1990) Let's razor Ockham's Razor. In his (1994) *From a biological point of view: essays in evolutionary philosophy*, 136-157. Cambridge: Cambridge University Press.

(1994) Prospects for an evolutionary ethics. In his (1994) *From a biological point of view: essays in evolutionary philosophy*, 93-113. Cambridge: Cambridge University Press.

(1995) Natural selection and distributive explanation: a reply to Neander. *British Journal for the Philosophy of Science* 46, 384-397.

(2012) Remarkable facts. *The Boston Review* November 07, 2012. Last accessed 22.07.2013: <http://www.bostonreview.net/books-ideas/remarkable-facts>

Sober, Elliott & Wilson, David Sloan (1998) *Unto others: the evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Sommers, Tamler & Rosenberg, Alex (2003) Darwin's nihilistic idea: evolution and the meaninglessness of life. *Biology and Philosophy* 18, 653-668.

Sosa, Ernest (1999) How to defeat opposition to Moore. *Philosophical Perspectives* 13, 137-149.

Spelke, Elizabeth S. & Kinzler, Katherine D. (2007) Core knowledge. *Developmental Science* 10, 89-96.

Sperber, Dan (1996) *Explaining culture: a naturalistic approach*. Oxford: Blackwell.

Sprevak, Mark (2009) Extended cognition and functionalism. *Journal of Philosophy* 106, 503-527.

Sripada, Chandra Sekhar (2008) Nativism and moral psychology: three models of the innate structure that shapes the contents of moral norms. In Walter Sinnott-Armstrong, ed. (2008) *Moral psychology, vol. 1: the evolution of morality*, 319-344. Cambridge, MA: MIT Press.

Stalnaker, Robert (1968) A theory of conditionals. In Nicholas Rescher, ed. (1968) *Studies in logical theory*, 98-112. Oxford: Blackwell.

Statman, Daniel (1991) Moral and epistemic luck. *Ratio* 4, 146–156.

Stegmann, Ulrich E. (2010) What can natural selection explain? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 41, 61–66.

Stephens, Christopher L. (2001) When is it selectively advantageous to have true beliefs? Sandwiching the better safe than sorry argument. *Philosophical Studies* 105, 161-189.

Stewart-Williams, Steve (2005) Innate ideas as a naturalistic source of metaphysical knowledge. *Biology and Philosophy* 20, 791-814.

Stout, Jeffrey (1988) *Ethics after Babel: the language of morals and their discontents.* Cambridge: James Clarke & Co.

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

(2008a) Reply to Copp: naturalism, normativity, and the varieties of realism worth worrying about. *Philosophical Issues* 18, 207-228.

(2008b) Constructivism about reasons. In Russ Shafer-Landau, ed. (2008) *Oxford studies in meta-ethics, vol. 3*, 207-246. Oxford: Oxford University Press.

(2011) Mind-independence without mystery: why quasi-realists can't have it both ways. In Russ Shafer-Landau, ed. (2011) *Oxford studies in meta-ethics, vol. 6*, 1-32. Oxford: Oxford University Press.

Sturgeon, Nicholas L. (1985) Moral explanations. In Geoffrey Sayre-McCord, ed. (1988) *Essays on moral realism*, 229-255. Ithaca, NY: Cornell University Press.

(1986) Harman on moral explanations of natural facts. *Southern Journal of Philosophy* 24, suppl. 43-55.

(2006) Moral explanations defended. In James Dreier, ed. (2006) *Contemporary debates in moral theory*, 241-262. Oxford: Blackwell.

Sugden, Robert (1986) *The economics of rights, co-operation and welfare.* Oxford: Blackwell.

Swain, Stacey, Joshua Alexander, & Jonathan Weinberg (2008) The instability of philosophical intuitions: running hot and cold on Truetemp. *Philosophy and Phenomenological Research* 76, 138-155.

Swinburne, Richard (1990) Argument from the fine-tuning of the universe. In John Leslie, ed. (1990) *Physical cosmology and philosophy*, 154-173. London: Macmillan.

(1997) *Simplicity as evidence for truth.* Milwaukee, WI: Marquette University Press.

Tajfel, Henri (1970) Experiments in intergroup discrimination. *Scientific American* 223, 96-102.

Tersman, Folke (2006) *Moral disagreement.* Cambridge: Cambridge University Press.

(2008) The reliability of moral intuitions: a challenge from neuroscience. *Australasian Journal of Philosophy* 86, 389-405.

Thomson, Judith Jarvis (1971) A defense of abortion. *Philosophy and Public Affairs* 1, 47-66.

(1985) The trolley problem. *The Yale Law Journal* 94, 1395-1415.

(1996) Moral objectivity. In Gilbert Harman & Judith Jarvis Thomson (1996) *Moral relativism and moral objectivity*, 65-154. Oxford: Blackwell.

Tinbergen, Niko, G. J. Broukhuysen, F. Feekes, J.C.W. Houghton, H. Kruuk, & E. Szulc (1962) Egg shell removal by the black-headed gull, *Larus ridibundus L.*; A behaviour component of camouflage. *Behaviour* 19, 74-117.

Tinbergen, Niko (1963) On the aims and methods of ethology. *Zeitschrift für Tierpsychologie* 20, 410-433.

Tomasello, Michael (1999) *The cultural origins of human cognition.* Cambridge, MA: Harvard University Press.

Tooby, John & Cosmides, Leda (1990) The past explains the present: emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology* 11, 375-424.

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

(2006) Reciprocal altruism: 30 years later. In Peter Kappeler & Carel Van Schaik, eds. (2006) *Cooperation in primates and humans: mechanisms and evolution*, 67-84. Heidelberg: Springer.

Turiel, Elliott (1983) *The development of social knowledge: morality & convention.* Cambridge: Cambridge, University Press.

(2006) The development of morality. In William Damon, Richard M. Lerner, & Nancy Eisenberg, eds. (2006). *Handbook of child psychology, 6th ed., vol. 3: social, emotional, and personality development*, 789-857. Hoboken, NJ: John Wiley & Sons.

Unger, Peter (1968) An analysis of factual knowledge. *The Journal of Philosophy* 65, 157-170.

Valdesolo, Piercarlo & DeSteno, David (2006) Manipulations of emotional context shape moral judgment. *Psychological Science* 17, 476-477.

Van Cleve, James (1999) *Problems from Kant.* Oxford: Oxford University Press.

van Fraassen, Bas (1980) *The scientific image.* Oxford: Oxford University Press.

(1984) Belief and the will. *Journal of Philosophy* 81, 235-56.

(1989) *Laws and symmetry.* Oxford: Oxford University Press.

Van Vugt, Mark & Iredale, Wendy (2013) Men behaving nicely: public goods as peacock tails. *British Journal of Psychology* 1, 3-13.

Vehrencamp, Sandra L. (1983) A model for the evolution of despotic versus egalitarian societies. *Animal Behavior* 31, 667-682.

Vogel, Jonathan (1987) Tracking, closure, and inductive knowledge. In Stephen Luper-Foy, ed. (1987) *The possibility of knowledge: Nozick and his critics*, 197–215. Totowa, NJ: Rowman and Littlefield.

(1990) Cartesian skepticism and inference to the best explanation. *Journal of Philosophy* 87, 658-666.

von Mises, Richard (1957) *Probability, statistics and truth*, transl. Geiringer. London: George Allen & Unwin.

Vonk, Jennifer, Sarah F. Brosnan, Joan B. Silk, Joseph Henrich, Amanda S. Richardson, Susan P. Lambeth, Steven J. Schapiro, & Daniel J. Povinelli (2008) Chimpanzees do

not take advantage of very low cost opportunities to deliver food to unrelated group members. *Animal Behaviour* 75, 1757–1770.

Watts, David P. (2002) Reciprocity and interchange in the social relations of wild male chimpanzees. *Behaviour* 139, 343-370.

Wedgwood, Ralph (2007) *The nature of normativity*. Oxford: Oxford University Press.

(2010) The moral evil demons. In Richard Feldman & Ted A. Warfield, eds. (2010) *Disagreement*, 216-246. Oxford: Oxford University Press.

Weinberg, Jonathan M., Chad Gonnerman, Cameron Buckner, & Joshua Alexander (2010) Are philosophers expert intuiters? *Philosophical Psychology* 23, 331-355.

Weinberg, Jonathan M., Shaun Nichols, & Stephen Stich (2001) Normativity and epistemic intuitions. *Philosophical Topics* 29, 429-460.

West, Stuart A., Claire El Mouden, & Andy Gardner (2011) Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior* 32, 231-262.

Westen, Drew (2007) *The political brain: the role of emotion in deciding the fate of the nation*. New York, NY: Public Affairs.

Westermarck, Edward (1891/1922) *History of human marriage, vol. 2*. New York: Allerton.

Wheatley, Thalia & Haidt, Jonathan (2005) Hypnotic disgust makes moral judgments more severe. *Psychological Science* 16, 780-784.

White, Roger (2000) Fine-tuning and multiple universes. *Noûs* 34, 260–276.

(2006) Problems for dogmatism. *Philosophical Studies* 131, 525-557.

(2009) Evidential symmetry and mushy credence. In Tamar Gendler & John Hawthorne, eds. (2009) *Oxford studies in epistemology, vol. 3*, 161-186.

(2010) You just believe that because… *Philosophical Perspectives* 24, 573-615.

Whiten, A., J. Goodall, W. C. McGrew, T. Nishida, V. Reynolds, Y. Sugiyama, C. E. G.

Tutin, R. W Wrangham, & C. Boesch (1999) Culture in chimpanzees. *Nature* 399, 682-685.

Wielenberg, Erik J. (2010) On the evolutionary debunking of morality. *Ethics* 120, 441-464.

Wiessner, Polly (1996) Levelling the hunter: constraints on the status quest in foraging societies. In Polly Wiessner & Wulf Schiefenhovel, eds. (1996) *Food and the status quest: an interdisciplinary perspective*, 171-191. Oxford: Berghahn Press.

(2005) Norm enforcement among the Ju/'hoansi Bushmen: a case for strong recriprocity? *Human Nature* 16, 115-145.

Wilkins, John S., & Griffiths, Paul E. (2013) Evolutionary debunking arguments in three domains: fact, value, and religion. In Greg Dawes & James Mclaurin, eds. (2013) *A new science of religion*, 133-146. London: Routledge.

Williams, Bernard (1973) A critique of utilitarianism. In J. J. C. Smart & Bernard Williams (1973) *Utilitarianism: for and against*. Cambridge: Cambridge University Press.

(1986) *Ethics and the limits of philosophy*. Cambridge, MA: Harvard University Press.

Williams, George C. (1966) *Adaptation and natural selection: a critique of some current evolutionary thought*. Princeton, NJ: Princeton University Press.

Williamson, Jon (1999) Countable Additivity and subjective probability. *British Journal for the Philosophy of Science* 50, 401-416.

Williamson, Timothy (1986) The contingent *a priori*: has it anything to do with indexicals? *Analysis* 46, 113 - 117.

(1988) The contingent *a priori*: a reply. *Analysis* 48, 218 - 221.

(2000) *Knowledge and its limits*. Oxford: Oxford University Press.

(2006) Conceptual truth. *Aristotelian Society Supplementary Volume* 80, 1–41.

(2007) *The philosophy of philosophy*. Oxford: Blackwell.

(2011) Philosophical expertise and the burden of proof. *Metaphilosophy* 42, 215-229.

Wilson, David Sloan & Wilson, Edward O. (2007) Rethinking the theoretical foundations of sociobiology. *The Quarterly Review of Biology* 82, 327-348.

Wilson, Edward O. (1975) *Sociobiology: the new synthesis*. Cambridge, MA: Harvard University Press.

(2004) *Human nature, revised ed.* Cambridge, MA: Harvard University Press.

Wolff, Robert Paul (1998) *In defense of anarchism, 2nd ed.* London: University of California Press.

Wong, David (1984) *Moral relativity*. Berkeley, CA: University of California Press.

(2006) *Natural moralities: a defence of pluralistic relativism*. Oxford: Oxford University Press.

Wright, Crispin (1992) *Truth and objectivity*. Cambridge, MA: Harvard University Press.

(1995) Truth in ethics. *Ratio* 8, 209-226.

(2004) Warrant for nothing (and foundations for free)? *Aristotelian Society Supplementary Volume* 78, 167–212.

Wright, Larry (1973) Functions. *The Philosophical Review* 82, 139-168.

Yagisawa, Takashi (1988) Beyond possible worlds. *Philosophical Studies* 53, 175-204.

(2010) *Worlds and individuals, possible and otherwise*. Oxford: Oxford University Press.

Zagzebski, Linda (1999) What is knowledge? In John Greco & Ernest Sosa, eds. (1999) *The Blackwell guide to epistemology*, 92-116. Oxford: Blackwell.

Zahavi, Amotz (1975) Mate selection - A selection for a handicap. *Journal of Theoretical Biology* 53, 205– 213.

(1995) Altruism as a handicap - The limitations of kin selection and reciprocity. *Avian Biology* 26, 1-3.

Zahn-Waxler, Carolyn, Marian Radke-Yarrow, Elizabeth Wagner, & Michael Chapman (1992) Development of concern for others. *Developmental Psychology* 28, 126-136.

Zangwill, Nicholas (2006) Moral epistemology and the because constraint. In James Dreier, ed. (2006) *Contemporary debates in moral theory*, 263-281. Oxford: Blackwell.

Zimmerman, David (1985) Moral realism and explanatory necessity. In David Copp and David Zimmerman, eds. (1985) *Morality, reason, and truth*, 79-103. Totowa, NJ: Rowman and Allanheld.

Zusne, Leonard & Jones, Warren H. (1989) *Anomalistic psychology: a study of magical thinking, 2nd ed.* Hillsdale, New Jersey: Lawrence Erlbaum.