Moral Skepticism

Moral skepticism is at present a vibrant topic of philosophical inquiry. Particularly since the turn of the millennium, the debates between moral skeptics of various stripes and their opponents have gained renewed force not only by taking account of innovative ideas in moral philosophy, but also by drawing on novel positions in epistemology, metaphysics, and philosophy of language as well as on recent findings in empirical sciences. As a result, new arguments for and against moral skepticism have been devised, while the traditional ones have been reexamined. This collection of original essays will advance the ongoing debates about various forms of moral skepticism by discussing such topics as error theory, disagreement, constructivism, non-naturalism, expressivism, fictionalism, and evolutionary debunking arguments. It will be a valuable resource for academics and advanced students working in metaethics and moral philosophy more generally.

Diego E. Machuca is Associate Researcher in Philosophy at CONICET (Argentina) and co-editor of the *International Journal for the Study of Skepticism*. He is also sole editor of *Pyrrhonism in Ancient, Modern, and Contemporary Philosophy* (2011), *New Essays on Ancient Pyrrhonism* (2011), and *Disagreement and Skepticism* (Routledge, 2013), and co-editor of two other volumes on skepticism.

Routledge Studies in Ethics and Moral Theory

For a full list of titles in this series, please visit www.routledge.com

Moral Skepticism

New Essays

Edited by Diego E. Machuca

Routledge
Taylor & Francis Group

NEW YORK AND LONDON

Contents

Acknowledgments

1 Moral Skepticism
An Introduction and Overview

Diego E. Machuca

1. Introduction

Introductory chapters of edited collections are hardly ever read, and are often rather brief. Despite the double risk of writing something that might not be perused and in a manner in which it is not often done, I will here not only present the essays that make up this volume but also offer an extensive critical overview of moral skepticism with the hope that it will turn out to be useful particularly to the uninitiated reader. I will first provide a taxonomy of varieties of moral skepticism, then discuss the main arguments advanced in their favor, and finally summarize the ten essays here collected, which deal with one or more of those skeptical stances and arguments. But before getting down to business, let me clarify the purpose of the present volume and say something about the peculiarity of its topic.

The aim has not been to put together a collection of essays that would jointly provide a comprehensive treatment of moral skepticism in the manner of a companion or a handbook. Rather, given the fertility of metaethical discussions of skepticism over the past fifteen years, it seemed timely to edit a volume of new research papers that would reexamine old issues in a fresh light, motivate further exploration of them, and introduce novel views. To the best of my knowledge, this is the first collection entirely devoted to exploring distinct varieties of moral skepticism.

An intriguing aspect of metaethics is that it is one of the few areas of philosophy—the others being philosophy of religion and philosophy of action—in which at present one finds quite a number of real skeptics, of one or another kind. In general, philosophers deem the importance of skepticism to be merely methodological, i.e., they regard skeptical arguments as useful tools for their inquiries. For instance, even though epistemologists think that skeptical arguments cannot be dismissed out of hand, most of them take it as plain that their conclusions are false and hence that there are mistakes somewhere in their premises. Careful analysis aimed at discovering the mistakes is considered philosophically useful and rewarding insofar as it allows us to get rid of the erroneous epistemological views expressed by the mistaken premises, and insofar as it allows us to acquire a deeper

understanding of the nature and scope of knowledge and justified belief in general or in particular areas. What are the reasons for there being quite a number of real skeptics in areas like metaethics, philosophy of religion, or philosophy of action? They are perhaps the fact that their subject matters are highly controversial and the fact that life can go on even if one denies, or suspends judgment about, the objectivity of morality, the existence of God, or the existence of free will.[1] Although purely epistemological matters are highly controversial as well, at least the great majority of epistemologists agree on the possibility of knowledge or justified belief in general, differing on how best to characterize their nature and scope. Also, denying, or suspending judgment about, the possibility of knowledge or justified belief in general would have many more damaging implications for our lives for the simple reason that it would target our beliefs as a whole. Whereas moral skepticism or free will skepticism might render certain kinds of action impossible—such as moral, responsible, or free action—radical epistemological skepticism would render action *tout court* impossible—"or so it is claimed," a Pyrrhonian skeptic would immediately add.

2.   The Multiple Faces of Moral Skepticism

It is important to make clear the range of views that are taken as varieties of moral skepticism in the present volume. The essays here collected deal not only with skepticism about moral knowledge or moral justification, but also with skepticism about moral reality. In other words, they deal with both epistemological and ontological forms of moral skepticism. Moral anti-realism (or irrealism, as some prefer to call it) is therefore treated as a variety of moral skepticism. This remark will strike metaethicists as obvious and hence unnecessary, but I make it to respond to an objection sometimes voiced, most particularly by epistemologists. The objectors argue that it is a surprising mistake to consider moral anti-realism a form of moral skepticism inasmuch as it does not target the possibility of moral knowledge or the epistemic justification of moral beliefs.[2] This objection reveals more the background of its proponents than the illegitimacy of the label. Note, first, that it is common among metaethicists to regard moral anti-realism as a kind of moral skepticism. Two examples might suffice. J. L. Mackie defended a position according to which first-order moral judgments are all false because the objective moral values, prescriptions, qualities, or relations they purport to describe do not exist. He called his position "moral skepticism" (Mackie 1977: 16–18, 35, 48–49; cf. 1946: 80–81, 83, 85, 90). And Walter Sinnott-Armstrong, in his comprehensive taxonomy of varieties of moral skepticism, includes what he calls "skepticism about moral reality" (2006: 12). A common practice may of course be mistaken. It is legitimate, however, to deem the various forms of moral anti-realism as skeptical for two interrelated reasons. First, moral anti-realism can be taken to target also moral knowledge or justified moral belief inasmuch as it claims that there

are no moral facts, properties, or relations to be known or about which to hold justified beliefs. Second, moral anti-realism calls into question most people's beliefs about morality by claiming either that all of our first-order moral judgments are false because there are no objective moral facts, properties, or relations; or that they are all neither true nor false because the moral facts, properties, or relations they presuppose do not exist; or that moral judgments are actually expressions of non-cognitive attitudes and not assertions about alleged mind-independent moral facts, properties, or relations. (More on these distinct views in a moment.) Hence, though some might be reluctant to regard moral anti-realism as a form of skepticism, not only is it a fact that it is commonly regarded that way among metaethicists, but there are also good reasons for so doing.

How to define moral anti-realism? The answer of course depends on how one conceives of moral realism. In the metaethical literature, a common distinction is that between minimal (or minimalist) and non-minimal (or non-minimalist) moral realism. According to the former, moral propositions are truth-apt and some of them are true. On this conception of moral realism, moral relativism could be considered a form of moral realism inasmuch as it affirms that certain moral propositions are true relative to a given framework. Moral constructivism, too, could be deemed to be a type of moral realism inasmuch as it maintains that certain moral propositions are true if they are those to which agents would agree, were they to engage in an idealized process of rational deliberation. Moreover, those versions of moral non-cognitivism that endorse a deflationary account of truth could also be regarded as forms of moral realism inasmuch as they accept that some moral sentences are true: to say that the sentence "Stealing is wrong" is true is just to say that stealing is wrong.[3] Non-minimal moral realism maintains, in addition, that some moral propositions are true by virtue of something in the world, namely, the objective or mind-independent moral facts or properties that those judgments track. This second form of moral realism can be either naturalistic or robust: roughly put, whereas naturalistic moral realism contends that moral facts and properties are either identical with or reducible to natural ones, robust moral realism claims that moral facts and properties are non-natural or irreducibly moral and hence causally inert.[4] In line with J. L. Mackie and Richard Joyce, I think that moral thought is inherently committed to the idea that objective moral facts or properties are intrinsically prescriptive—they categorically demand or require that people act in certain ways irrespective of their desires, aims, or interests—and that moral naturalism fails to account for such intrinsic prescriptivity (or inescapable authority or irreducible normativity).[5] If so, then a commitment to the existence of objective moral facts or properties that are intrinsically prescriptive characterizes not only the position of those *philosophers* who are robust moral realists, but also *ordinary* moral thought and discourse. Somewhat less contentiously, it seems that ordinary people are typically non-minimal moral realists of some sort, even though at least

the great majority of them are of course unable to articulate their position the way metaethicists do.[6]

Given these considerations, a possible general formulation of the moral anti-realist's ontological skepticism is the following:

Ontological Moral Skepticism

There are no objective or mind-independent moral facts or properties.

This formulation has the advantage of encompassing all those views that reject the ontological commitment of non-minimal moral realism. The two main views to be mentioned are moral error theory and moral non-cognitivism (also known as "non-descriptivism"). Since I will focus on the former, let us start with the latter, whose standard version could be formulated thus:

Moral Non-Cognitivism

Moral judgments are not truth-apt because they are expressions of non-cognitive attitudes or states (such as emotions or commands), not assertions that convey beliefs about alleged objective moral facts or properties.

This view is a form of ontological moral skepticism because it maintains not only that moral judgments are not descriptions of objective moral facts or properties, but also that these facts or properties do not exist. Insofar as they are not statements of matters of fact, moral judgments are radically different from non-moral ones, which also explains the intimate connection between moral thought and motivation. As we will see, the claim that moral judgments are not assertions that convey beliefs, despite their being usually expressed in the indicative mood, is what distinguishes moral non-cognitivism from moral error theory. It should be noted, however, that there is a form of moral non-cognitivism that is milder inasmuch as it holds that, though moral judgments are primarily expressions of non-cognitive attitudes, they also express beliefs. This view is commonly dubbed a "hybrid" form of moral non-cognitivism.[7]

A possible formulation of the other main type of ontological moral skepticism is this:

Moral Error Theory

First-order moral judgments are truth-apt because they are assertions that attribute moral properties to objects, but they are all false because such properties do not exist or are not instantiated.[8]

This is an *error* theory precisely because it claims that, in making first-order moral judgments, we misdescribe or misrepresent the world inasmuch as it

does not contain the items posited, implied, or presupposed by those judgments. Such a formulation of moral error theory corresponds to the way it is typically understood (see esp. Mackie 1977: 35, 48–49). There are also nonstandard versions, one of which can be mentioned here because its departure from the standard one is not significant.[9] It maintains that, given that there is a referential or presupposition failure in moral judgments inasmuch as they refer to, or presuppose the existence of, objective moral facts, properties, or relations that nonetheless do not exist, those judgments are neither true nor false (see esp. Joyce 2001: 6–9). So a slightly better formulation of moral error theory would say, not that first-order moral judgments are all false, but that they are all untrue, which may be understood either in the sense that they are all false or in the sense that they are all neither true nor false.

It is worth noting that Mark Eli Kalderon (2005: 105–106, 144–145) claims that the standard formulation of moral error theory should be revised so as to also include moral agnosticism: "Competent speakers should not believe [moral] propositions expressed by the target [moral] sentences that they accept either because they are false or because they are unjustified" (2005: 106).[10] The problem with this revised formulation is that it creates confusion inasmuch as a moral agnostic who remarks that the available evidence justifies neither moral realism nor moral anti-realism refrains, for that very reason, from affirming that there is a fundamental *error* in moral discourse in that this discourse is committed to an *erroneous* picture of the world. Saying that moral beliefs are unjustified is clearly different from saying that they are erroneous.

An error theory is a theory about a given discourse, not a cluster of items, and defining moral error theory as the denial of the existence of objective moral facts, properties, or relations does not allow one to distinguish it from other types of moral anti-realism (cf. Joyce & Kirchin 2010: xii). What sets it apart is the view that moral judgments are assertions that express beliefs. But this should not make us lose sight of the fact that it is the ontological element of moral error theory that leads its proponent to affirm that morality has been undermined or debunked. Note, in this regard, that Mackie explicitly presents his skeptical stance as an ontological thesis and remarks that its linguistic aspect is a corollary of the ontological aspect, which is the central one:

> [W]hat I have called moral scepticism is an ontological thesis, not a linguistic or conceptual one. It is not, like the other doctrine often called moral subjectivism, a view about the meanings of moral statements. Again, no doubt, if it is to be at all plausible, it will have to give some account of their meanings. . . . But this too will be a development of the theory, not its core.
>
> (1977: 18)

These remarks make perfect sense in light of the fact that Mackie opens the first chapter of *Ethics* with the assertion "There are no objective moral values" (1977: 15).

The other main variety of moral skepticism is epistemological in nature:

Epistemological Moral Skepticism

We do not possess moral knowledge or epistemically justified moral beliefs.

The disjunction in this formulation is due to the fact that epistemological moral skepticism may either target only moral knowledge or be broader in scope and target epistemically justified moral belief. The formulation also attempts to capture two other distinct stances: one extreme that denies the very possibility of moral knowledge or of epistemically justified moral belief, the other more cautious that recommends adopting an agnostic attitude:

Nihilistic Epistemological Moral Skepticism

Moral knowledge is impossible or no moral belief is ever epistemically justified.

Pyrrhonian Moral Skepticism

One must to suspend judgment about whether moral knowledge is possible and about whether any moral belief is epistemically justified.

It is nihilistic epistemological moral skepticism that has been more commonly discussed in the contemporary metaethical literature. However, Pyrrhonian skepticism has slowly but increasingly been taken into consideration, perhaps due to the fact that it has, for some time now, been the focus of much attention in epistemology. I will here limit myself to making three sets of remarks about both forms of epistemological moral skepticism, the first two concerning the nihilistic variety.

First, David Enoch (2011: 4–5 with n. 7) claims that robust normative realism (and hence robust moral realism) is compatible with what he calls "epistemological normative skepticism," which claims that no normative belief (and hence no moral belief) is epistemically justified or amounts to knowledge. Though Enoch is right that both views are in principle compatible, I take it that any consistent epistemological moral skepticism must be broad in scope, i.e., that it must target the epistemic credentials not only of first-order moral beliefs but also of the second-order belief that there are objective and irreducibly moral facts. Any epistemological moral skeptic worth his salt will ask how it is that the robust moral realist has come to have cognitive access to the existence of such facts.

Second, Joyce, who has defended both a moral error theory (Joyce 2001) and a nihilistic skepticism about moral justification (Joyce 2006), maintains that these two skeptical stances may or may not be held together:

One might endorse an error theory while maintaining that people are justified in their moral beliefs, or alternatively endorse an error theory

while adding that all people's moral beliefs lack justification. Similarly, the claim that moral beliefs lack justification may combine with the view that they are all false, but is also consistent with the possibility that moral beliefs are not only true but objectively true . . . . [J]ustification skepticism is compatible with a realist stance.

(2016b: 1–2)

I disagree with Joyce for two reasons. First, a moral error theorist can consistently maintain that *others* may be, in some sense, epistemically justified in their moral beliefs, but not that *he himself* is. Take the distinction between what are sometimes called the "subjective" and "objective" components of justification: the former refers to whether the subject has responsibly formed a given belief (e.g., there is no evidence of its falsity of which he is aware), while the latter refers to whether the belief has been reliably formed (e.g., the belief tracks the truth). The moral error theorist may realize that, although the total body of the available evidence indicates that all first-order moral beliefs are false, ordinary people are subjectively justified in holding first-order moral beliefs because they are not aware of the evidence of their falsity. By contrast, it does not seem possible for the moral error theorist not to be skeptical about moral justification: if he believes that there are no objective moral facts or properties and, hence, that all first-order moral beliefs are false, then he must conclude that no such beliefs are epistemically justified, either objectively or subjectively. He is fully aware of the undefeated reasons against his former first-order moral beliefs and believes to know that these beliefs do not track the truth. Second, whereas it is not possible for a moral error theorist not to be a skeptic about moral justification, it is possible for a skeptic about moral justification not to be a moral error theorist or some other kind of moral anti-realist, since for all he knows there might be moral facts or properties out there in the world. But note that this is different from claiming that skepticism about moral justification is compatible with the adoption of a non-minimal moral realist view. A non-minimal moral realist could, in principle, affirm that there are objective moral facts or properties, but deny that our moral beliefs are ever justified or that they amount to knowledge. But the epistemological moral skeptic would ask how the non-minimal moral realist can know or justifiably believe that there is an objective moral reality: the latter would have to explain how he can have such metaethical knowledge or such justified metaethical belief, but lack first-order moral knowledge or justified first-order moral beliefs. Why do we suffer from a serious cognitive limitation in one case, but not in the other?

Third, it must be remarked that a Pyrrhonian moral skeptic suspends judgment not only about whether anyone knows or justifiably believes that something is morally right or wrong, but also about the epistemic credentials of the various metaethical views. The reason is that both first-order and second-order disagreements have, at least thus far, struck him as unresolvable. The second-order disagreements include not only those amongst the various

realist views, but also those between moral realism and the non-Pyrrhonian moral skeptical stances. For the Pyrrhonian skeptic suspends judgment with regard to the metaethical debate about the existence of objective moral facts or properties and with regard to the metaethical debate about the possibility of moral knowledge or epistemically justified moral belief. This is why he says that we do not have moral knowledge or epistemically justified moral beliefs, but refrains from denying that we will ever do.

How could moral skepticism be defined in a way that encompassed the various stances that have been distinguished in the present section? Perhaps as the view that undermines or debunks morality by attacking its ontological foundation and/or the epistemic credentials of moral belief. I proceed now to present and discuss—unfortunately, but inevitably, in an incomplete manner—the four main arguments for moral skepticism that have been advanced in the literature.

3.   Arguments for Moral Skepticism

A good way to start the discussion of the main arguments in favor of moral skepticism may be by quoting a passage in which Mackie summarizes the five points that support his skeptical position:

> The considerations that favour moral scepticism are: first, the relativity or variability of some important starting points of moral thinking and their apparent dependence on actual ways of life; secondly, the metaphysical peculiarity of the supposed objective values, in that they would have to be intrinsically action-guiding and motivating; thirdly, the problem of how such values could be consequential or supervenient upon natural features; fourthly, the corresponding epistemological difficulty of accounting for our knowledge of value entities or features and of their links with the features on which they would be consequential; fifthly, the possibility of explaining, in terms of several different patterns of objectification, traces of which remain in moral language and moral concepts, how even if there were no such objective values people not only might have come to suppose that there are but also might persist firmly in that belief.
>
> (1977: 48–49)

The first four considerations refer to two skeptical arguments. The first is a reference to the argument that Mackie calls "the argument from relativity," but that is more accurately viewed as an argument from disagreement. Considerations two to four are parts of the same argument, the argument from queerness, or rather refer to three different versions of the argument. The passage does not mention the other two main arguments advanced in the literature: the argument from the best explanation and the argument from evolution. However, Mackie's arguments from disagreement and queerness can, as we will see, be constructed as versions of the argument

from the best explanation, and at certain points he mentions evolution as a possible account of the origin of morality. Finally, the fifth consideration concerns the thesis of moral objectification or moral projectivism, which I will also briefly discuss after dealing with the above four arguments. The reason is that it has been, or could be, proposed as a (key) supplement to some of those arguments insofar as it would explain why we systematically make the moral error or hold beliefs that are epistemically unjustified.

The order of exposition will be as follows: the argument from the best explanation, the argument from disagreement, the argument from queerness, the argument from evolution, and the objectification thesis. Given the recent explosion of interest in the evolutionary debunking of morality, I will focus primarily on the argument from evolution. Besides the prominent place they occupy in the literature on moral skepticism, the four arguments and the objectification thesis will be mentioned or discussed in several of the essays of the present volume.[11]

Before proceeding, it is important to briefly explain a now widely accepted distinction in epistemology between *rebutting* and *undercutting defeaters*, which goes back to Pollock (1986). A rebutting defeater for a proposition p is counterevidence for p that is stronger than one's original evidence for p. By contrast, an undercutting defeater for p is not evidence that p is false, but evidence that undermines the connection between p and one's original evidence for p by showing, for example, that the source of the belief that p is unreliable. In one case, one has evidence that p is false; in the other, one has evidence that one does not have sufficient reason to believe that p is true. Thus, whereas a rebutting defeater is a reason for believing the negation of p, an undercutting defeater is a reason for no longer believing p. The distinction is crucial because establishing that the available evidence does not actually justify a given claim does not tell us anything about its truth or falsity. With this distinction in place, it should be noted that, while some of the four arguments to be discussed provide a rebutting defeater for our moral beliefs, others provide an undercutting defeater. In fact, as we will see, the arguments provide one or the other kind of defeater depending on the version of the argument that is advanced.

3.1.   *Argument From the Best Explanation*

The argument from the best explanation, discussed particularly by Harman (1977: 7–10, 13, 130–132), claims that there are no moral facts or properties because they do not figure in the best explanation of why we have moral beliefs or make moral judgments. Here is a possible formulation of the argument:

1.   Our having moral beliefs is best explained by certain psychological and socio-cultural facts about us, not by there being moral facts.
2.   If moral facts are explanatorily redundant, then they do not exist.

Therefore:

3.   There are no moral facts.

The second premise can be interpreted as expressing a principle of parsimony, according to which one should not unnecessarily multiply entities: if a kind of entity is not necessary for explaining a given phenomenon, one should not affirm or accept its existence; moreover, one should deny that it exists. That explanatory redundancy or dispensability suffices by itself to assert that something does not exist is no doubt questionable. It could be argued that the second premise should instead be couched in epistemological terms: if certain facts are explanatorily redundant or dispensable, then one has no reason for believing in their existence. In this case, the conclusion of the argument would of course be epistemological as well. Interpreted in this way, the argument still raises a serious challenge to moral realists: they would have the burden of providing reasons for believing that there are moral facts or properties. Note that if the ontological version of argument from the best explanation were sound, it would provide a rebutting defeater for our moral beliefs: it would show that our moral beliefs are false because there are no moral facts or properties. If the epistemological version were sound, it would provide an undercutting defeater: it would show that the realist explanation of our having moral beliefs is not as good as we thought it was. Note also that, even if moral facts understood in a deflationary manner were not explanatorily redundant, the argument could still be formulated so as to target specifically non-minimal moral realism. For example, it could be argued that, although relative moral facts figure in the best explanation of our having moral beliefs, objective or mind-independent moral facts do not.

Let me finally observe that, as we will see in the next subsections, some of the other arguments for moral skepticism can be viewed as versions of the argument from the best explanation inasmuch as they include a premise that refers to the alleged best explanation of a given phenomenon.

3.2.   *Argument From Disagreement*

The argument from disagreement can be used to support both ontological and epistemological forms of moral skepticism, depending on the premises that accompany the one that refers to the existence of deep, persistent, and widespread disagreements about moral matters. Until recently, the versions of the argument most commonly discussed in contemporary metaethics were those purporting to establish an ontological conclusion. One such version, proposed by Mackie (1977: 37) to ground his moral error theory, includes a best-explanation premise: "the actual variations in the moral codes are more readily explained by the hypothesis that they reflect ways of life than by the hypothesis that they express perceptions, most of them seriously inadequate and badly distorted, of objective values" (1977: 37). The alternative realist explanation would require that we possess a moral faculty that is highly

unreliable inasmuch as our moral errors would not be sporadic and temporary, but recurrent and lasting (1946: 78). If there is no objective moral fact of the matter concerning any issue, then it is no mystery why people disagree deeply, persistently, and widely about what the objective moral fact of the matter is. The argument could be formulated as follows:

1. There exist deep, persistent, and widespread disagreements about moral matters.
2. Such moral disagreements are best explained as resulting from variations in ways of life or cultures or social conventions rather than from variations in perceptions of alleged objective moral facts or properties.
3. If objective moral facts or properties are explanatorily redundant, then they do not exist.

Therefore:

4. There are no objective moral facts or properties.

This argument can thus be deemed to be a combination of the argument from disagreement and the argument from the best explanation, and so it could perhaps be called 'the argument from the best explanation of disagreement'. Note that the role of disagreement is not irrelevant, since it is what raises the challenge to the moral realist: there is a phenomenon that needs to be accounted for. As observed at the beginning of the present section, Mackie called his version of the disagreement-based skeptical argument "the argument from relativity." Even though this argument exploits the existence of deep, persistent, and widespread moral disagreements, the relativity is seen in the fact that the plurality of conflicting moral beliefs is to be explained by these beliefs being relative to certain socio-cultural factors. Note that Mackie's ontological version of the argument, if sound, provides a rebutting defeater for our first-order moral beliefs. If premise 3 were instead couched in epistemological terms—i.e., if objective moral facts or properties are explanatorily redundant, then one has no reason for believing in their existence—then the argument would be epistemological in nature, thereby providing, if sound, an undercutting defeater for our first-order moral beliefs.

Another, probably stronger, epistemological version of the disagreement-based argument for moral skepticism emphasizes the (as yet) impossibility of coming up with a clear-cut and impartial way of resolving moral disputes, and concludes that conflicting moral beliefs are not epistemically justified or do not amount to knowledge either *per se* or up to this point. Here's a possible formulation of the argument, focusing on epistemic justification:

1. There exist deep, persistent, and widespread moral disagreements.
2. There is (as yet) no clear-cut and impartial way of epistemically resolving such disagreements.

Therefore:

3.   Conflicting moral beliefs are not epistemically justified *per se* or up to this point.

The (as yet) impossibility of epistemically resolving moral disagreements may be due to different reasons: the epistemic peerhood of the contending parties, the lack of an agreed-upon epistemic criterion, or the inability to meet the epistemic challenge posed by the so-called Agrippa's trilemma. The disjunction in the conclusion is introduced so as to include both nihilistic epistemological skepticism and Pyrrhonian skepticism. If sound, this argument provides an undercutting defeater for our first-order moral beliefs. Note that the argument can also be formulated so as to target metaethical views by pointing to the (as yet) impossibility of resolving the second-order disagreements between them, thereby concluding that they are not (as yet) epistemically justified. The epistemological skeptic who utilizes the argument in this way may call attention to the long-standing debates among champions of the various moral realist views. Or he may call attention to the equally plausible alternative explanations proposed by moral realists and anti-realists to explain first-order moral conflicts: these conflicts are due either (*i*) to the fact that only one of the rival parties has epistemic access to the moral fact of the matter while the others suffer from some cognitive deficiency or shortcoming, such as inferential error, ignorance of relevant evidence, or prejudice (as the moral realist contends); or (*ii*) to the fact that there is no such fact of the matter epistemic access to which would in principle make it possible to impartially adjudicate the disagreement (as the moral anti-realist maintains). The epistemological skeptic is agnostic about moral ontology.[12]

3.3.   *Argument From Queerness*

Another prominent argument for moral skepticism is that which focuses on the alleged queerness of certain aspects of morality. The main proponent of this argument and the one who expressed it most clearly was Mackie, who first regarded it as not being very strong or very plausible (1946: 78), but later on considered it as being even more important than the argument from disagreement (1977: 38). Here is his presentation of the argument:

> This has two parts, one metaphysical, the other epistemological. If there were objective values, then they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe. Correspondingly, if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else.
>
> (1977: 38; cf. 1982: 115, 238)

As we saw at the beginning of the present section, Mackie actually identifies three elements that are queer: the intrinsically action-guiding and motivating nature of the alleged objective moral facts, the supervenience of these facts upon natural ones, and the knowledge of both objective moral facts and their links with the natural ones upon which they supervene. The queer elements are in fact four, given that the first refers to two aspects of objective moral facts that are mysterious: they are action-guiding *and* motivating. Richard Garner (1990: 137, 142–144) and Richard Joyce (2001: 30–31) hold that the preferred reading of Mackie's argument is that according to which the real queerness of moral facts does not concern their alleged power to motivate, but their objective bindingness or prescriptivity or inescapability—their intrinsically action-guiding nature—and maintain that it is such a notion that makes it possible to construct a compelling argument for a moral error theory. Similarly, Jonas Olson (2014: chs. 5–6) distinguishes four versions of the argument from queerness that target supervenience, knowledge, motivation, and irreducible normativity, and claims that only the last one stands up to scrutiny. In their use of the argument to challenge moral realism, Terence Horgan and Mark Timmons (1992) prefer to "revive and rejuvenate" the version concerning supervenience.[13]

Like Mackie's version of the argument from disagreement, his queerness argument includes a best-explanation premise:

> [T]he intuition required might be the perception that wrongness is a higher order property belonging to certain natural properties; but what is this belonging of properties to other properties, and how can we discern it? How much simpler and more comprehensible the situation would be if we could replace the moral quality with some sort of subjective response which could be causally related to the detection of the natural features on which the supposed quality is said to be consequential.
>
> (1977: 41)

I interpret Mackie as saying not only that the supervenience of moral properties on natural ones is mysterious, both ontologically and epistemologically, but also that our moral judgments are better explained as subjective responses to natural properties—an explanation that implies nothing queer or bizarre— rather than as descriptions of moral properties that allegedly exist in the world. This better explanation may also include the objectification thesis (to be discussed in Subsection 3.5): our moral judgments are the result of affective attitudes that are caused by natural properties and that are projected onto the world. Thus, like the argument from the best explanation and his version of the argument from disagreement, Mackie's argument from queerness seems to include a premise that expresses a principle of parsimony: if a kind of entity is not indispensable for explaining a given phenomenon, one should deny its existence. A possible formulation of the argument is the following:

1. Accepting the objective truth of morality implies accepting the existence of ontologically and epistemologically queer entities, faculties, and relations.

2. Queer moral entities, faculties, or relations are not indispensable for explaining our moral beliefs.
3. If an item is explanatorily redundant, then it does not exist.

Therefore:

4. The queer entities, faculties, and relations posited by morality do not exist.

If sound, this argument provides a rebutting defeater for our first-order moral beliefs, because it shows that they are all false. But if premise 3 were deemed implausible and were therefore reformulated in epistemological terms—i.e., if an item is explanatorily dispensable, then one has no reason for believing in its existence—then the argument would be epistemological in nature and, if sound, would provide an undercutting defeater for our first-order moral beliefs by showing that they are epistemically unjustified.

3.4. *Argument From Evolution*

Drawing especially on the work of evolutionary biologists, some moral skeptics have argued that the most plausible account of the origin of morality is the one that appeals to evolution: natural selection has forged certain faculties or capacities devoted to moral judgment. In their view, the evolutionary account defeats our first-order moral beliefs because it does not require that morality be *true*, but only that it be evolutionarily advantageous to *believe* that it is true. Evolutionary debunking strategies of this sort have been deployed in a systematic way, particularly by Richard Joyce (2001: ch. 6; 2006; 2016c) and Sharon Street (2006; 2008). Joyce first appealed to the argument from evolution in his defense of a moral error theory, but later on used it to ground a skepticism about moral justification. Street employed the argument in her attack not merely on moral realism but on value realism in general. Although in the two articles in question she does not develop or defend it, she repeatedly mentions constructivism as the anti-realist view that sidesteps her evolutionary debunking argument against value realism.

The defense, interpretation, and criticism of various types of evolutionary arguments for moral skepticism have of late attracted a lot of attention, and in fact the study of 'the evolution of morality' constitutes a burgeoning area in metaethics. The thrust of such arguments is that biological evolution is aimed not at moral belief-forming processes that are reliable, but at moral belief-forming processes that are adaptive. In other words, the evolutionary function of those processes is not that of tracking the truth: their general success at matching or accurately representing alleged objective moral facts explains neither their emergence nor their persistence. Humans are therefore disposed to make moral judgments regardless of the evidence to which they are exposed, regardless of whether there are or are not objective moral

facts. Someone might object that, in order to be adaptive, such processes might be reliable, i.e., the moral judgments they form are evolutionarily useful—i.e., tend to promote survival and reproduction—because they are in general true. However, given that moral beliefs may well be adaptively useful even if they are not true, if what we know is only that evolution is aimed at moral belief-forming processes that are adaptive, then we do have here a defeater: even if some moral judgments are true, there is no reason for claiming that they are. This is the way in which evolutionary skeptical arguments are in general understood in the literature. Resuming the distinction between rebutting and undercutting defeaters discussed at the outset of the present section, the evolutionary account of the origin of our moral beliefs then provides an undercutting defeater for those beliefs: it does not show that they are false—for there might well be moral facts out there in the world—but rather that they were not formed in a reliable way because their source is not trustworthy, and hence that they are not epistemically justified. The resulting moral skepticism is therefore epistemological. However, as we will see, the evolutionary account has also been understood as providing a rebutting defeater for our moral beliefs: a reason for thinking that objective moral facts do not exist, and hence that such beliefs are false. The resulting moral skepticism is therefore ontological.

When appealed to in relation to a moral error theory, evolutionary debunking considerations are normally used as a supplement to arguments that purport to establish the error-theoretic conclusion in order to account, once the conclusion is accepted, for the systematic error we commit in making moral judgments. This seems to be the case of Mackie, who briefly appealed to evolution as an alternative explanation of the origin of our moral sentiments and dispositions (1977: 113–114, 124, 192, 229, 239). Although Mackie (1985: 154) claimed that morality can be seen as an outgrowth from genetically determined retributive tendencies that were favored by evolutionary selection,[14] he did not offer an elaborate evolutionary account of morality in the way Joyce (2001: ch. 6; 2006) has. The latter maintains that the origin of morality is to be found in the development of human cooperation: an individual is more reproductively fit if his sympathetic desires to help his family members are supplemented by a sense of inescapable requirement to favor them that strengthens his motivation to perform helpful actions. This was accomplished by providing people with the belief that such actions have objective moral qualities. Once a cognitive capacity to believe that it is inescapably required to help family members was in place, it was exploited by natural selection to regulate also helpful behavior towards non-kin individuals. It must be remarked that Joyce's view is not that every particular moral prescription can be evolutionarily explained, or that culture or the environment plays no role in determining moral beliefs. Rather, his view is that the tendency to use general moral categories and the belief that certain types of action bear objective moral properties are innate; that cultural influences can cause some of those actions to

stop being regarded as moral or immoral, or cause other types of action to start being so regarded; and that moral dispositions require environmental cues to become manifest. For reasons that will become clear at the end of this subsection, it is important to note that Joyce is at some points cautious regarding the status of his evolutionary account of morality. He presents the hypothesis that natural selection has led us to commit the fundamental moral error as a "plausible speculation" (2001: 135). Also, although he regards the evolutionary hypothesis as plausible, coherent, and testable, and as the best story of the origin of morality we have (2006: 134, 137, 139–140), and although he therefore answers the question "Is human morality innate?" in the affirmative, he remarks that "this is provisional and to a degree speculative, since the present evidence does not warrant answering the question in either a positive or a negative way with any confidence" (2006: 2). Finally, he observes that his evolutionary debunking argument "is conditional: It relies on an empirical premise concerning the evolution of morality which is yet to be established" (2016b: 9).

In his first treatment of the evolutionary account of morality, Joyce not only remarks that it complements the arguments for moral error theory, but he makes the stronger claim that "the fact that moral thinking is a naturally evolved trait has error theoretical implications" (2001: 137) or "provides evidence in favor of the error theory" (2001: 148). In his view, the

> innateness of moral judgments undermines these judgments being true for the simple reason that if we have evolved to make these judgments irrespective of their being true, then one could not hold that the judgments are *justified*. And if they are unjustified, then although they *could* be true, their truth is in doubt.
>
> (2001: 159)

But the fact that if we accept the evolutionary account, our moral beliefs are utterly unjustified, or we have no reason for thinking that they are true, or it is highly improbable or extremely unlikely that they are true, in no way establishes the ontological conclusion of moral error theory. Of course, the evolutionary account places the burden of proof on the non-minimal moral realist to provide us not only with a reason for believing that our moral beliefs are epistemically justified, but also with a reason for believing that there are objective moral facts or properties in the first place. Oddly enough, Joyce himself recognizes that the evolutionary account alone does not support an ontological conclusion, but rather an attitude of withholding of assent concerning the truth or falsity of moral judgments (2001: 160–168). In any case, in later works he explicitly remarks that one cannot argue for a moral error theory on the basis of evolutionary considerations, the correct skeptical conclusion being instead that all moral judgments are unjustified (Joyce 2006: ch. 6; 2016c; cf. 2016b: 8). Joyce's later evolutionary-debunking stance seems to vacillate between nihilistic and Pyrrhonian

epistemological skepticism: sometimes he seems to believe that moral beliefs are intrinsically unjustified or that they have been shown to be so for good, and sometimes to believe that they can be deemed to be unjustified on the basis of the evidence available up to this point. Joyce's epistemological version of the argument from evolution could be formulated as follows:

1. Our capacity to form first-order moral beliefs is an evolutionary adaptation produced by natural selection.
2. Biological evolution is not aimed at moral belief-forming processes that are reliable, i.e., processes whose function is to track the alleged moral truths.
3. Given 2, our having beliefs that objects possess moral properties is consistent with nothing ever possessing a moral property.

Therefore:

4. Our first-order moral beliefs are epistemically unjustified.

Street (2006) contends that evolutionary considerations pose a dilemma for realist theories of value (and hence for realist theories of moral value). The fact that the forces of natural selection have greatly shaped the content of our evaluative judgments raises the challenge to explain the relation between such evolutionary influences and the independent evaluative facts posited by the realist.[15] The first horn of the dilemma is the claim that there is no such relation, which results in an implausible skepticism: we would have to conclude that our evaluative judgments are contaminated by a distorting influence and hence that many or most of them are off the track. Although it is possible that "as a matter of sheer chance" our evaluative judgments accord with the allegedly independent evaluative facts, "this would require a fluke of luck that's not only extremely unlikely . . . but also astoundingly convenient to the realist" (2006: 122). In response, one could appeal to rational reflection as another major influence on the content of our evaluative judgments that corrects the distorting influence of evolutionary pressures on such judgments. Although Street does not discard such an influence, she claims that, since rational reflection must proceed by using evaluative judgments, one would be assessing evolutionarily distorted evaluative judgments by means of other evolutionarily distorted evaluative judgments (2006: 124).

The other horn of the dilemma is the claim that natural selection favored those ancestors who were able to grasp the independent evaluative truths, because tracking them was advantageous for survival and reproduction. But this account that presents itself as a scientific explanation is, in Street's view, inferior on scientific grounds to the one according to which the tendency to make certain kinds of evaluative judgments rather than others contributed to our ancestors' survival and reproduction, because those judgments

forged adaptive links between the circumstances in which our ancestors found themselves and their responses to such circumstances. This account is superior in terms of the usual criteria of scientific adequacy, for it is clearer, more parsimonious, and does a better job at illuminating the tendency in question (2006: 129–134). Once again, we see that a crucial premise in an argument against value realism is a best-explanation premise. With a focus on moral realism, Street's argument could perhaps be formulated thus:

1. The forces of natural selection have had an indirect tremendous influence on the content of our moral judgments.
2. The moral realist owes us an explanation of the relation between such an evolutionary influence and the independent moral facts he posits.
3. He can claim either that (3a) there is no relation or that (3b) there is such a relation.
4. If he claims that (3a), then he is forced either (4a) to embrace a far-fetched moral skepticism or (4b) to claim that an incredible coincidence took place.[16]
5. If he claims that (3b), then he must propose a tracking account, which is scientifically unacceptable (since the adaptive link account provides the *best explanation* of why our tendency to make certain kinds of moral judgments rather than others contributed to our ancestors' reproductive success).

Therefore:

6. Moral realism is false, i.e., there are no independent moral facts.

It is surprising that Street argues for an ontological conclusion regarding independent or objective moral facts on the basis of an evolutionary debunking argument. For it seems that evolutionary debunking arguments (and genealogical debunking arguments in general) can at most undermine the epistemic credentials of our substantive moral beliefs—i.e., can at most provide us with undercutting defeaters for those beliefs. Street's own evolutionary debunking argument establishes at most that we have no reason for affirming that our moral beliefs match alleged objective moral facts because the best explanation of our tendency to make certain moral judgments makes no appeal to them. Even though the moral realist then owes us a reason for affirming that such facts exist, the argument does not prove that they do not. Note that such epistemological moral skepticism corresponds to (4a), the skeptical conclusion that Street regards as implausible or far-fetched.

It is important to observe that Street, just like Joyce, expresses caution regarding the status of her evolutionary debunking argument. She claims:

> I attach a . . . caveat to my argument in this paper: if the evolutionary facts are roughly as I speculate, here is what might be said

philosophically. I try to rest my arguments on the least controversial, most well-founded evolutionary speculations possible. But they are speculations nonetheless.

<div align="right">(2006: 112)</div>

And then she adds:

[I]t must suffice to emphasize the hypothetical nature of my arguments, and to say that while I am skeptical of the details of the evolutionary picture I offer, I think its outlines are certain enough to make it well worth exploring the philosophical implications.

<div align="right">(2006: 113)[17]</div>

There have been several recent attempts to refute evolutionary debunking arguments, of which I would like to mention two. First, David Enoch (2011: ch. 7) claims that Street's Darwinian dilemma is a particular instance of what he calls "the epistemological challenge." Such a challenge consists in the demand of an explanation of the correlation between normative truths and our normative judgments or beliefs:

What explains this correlation? On a robustly realist view of normativity, it can't be that our normative judgments are causally or constitutively responsible for the normative truths, because the normative truths are supposed to be independent of our normative judgments. And given that (at least basic) normative truths are causally inert, they are not causally responsible for our normative beliefs.

<div align="right">(2011: 159)</div>

According to Enoch's robust normative realism, then, normative truths are not causally efficacious, and so the demand for an explanation is more pressing for it than for other forms of normative realism. Enoch offers "a third-factor explanation, or indeed a (Godless) pre-established-harmony type of explanation" (2011: 168). Assuming that survival or reproductive success is, not always or intrinsically, but somewhat or by and large good in the sense that it is an aim recommended by normative truths, then given that selective forces have shaped our normative beliefs so as to achieve survival or reproductive success, "our normative beliefs have developed to be at least somewhat in line with the normative truths" (2011: 168). More precisely:

Selective forces have causally shaped our normative beliefs; that survival is good is (non-causally but closely) related to many normative truths; and so that survival . . . is good explains the correlation between our normative beliefs and the normative truths.

<div align="right">(2011: 169–170)</div>

It must be noted that Enoch remarks that he has "no idea whether this explanation actually works (or whether the phenomenon it is supposed to explain is actually a real phenomenon)," and that all that is crucial for him "is that it *could* work, and that its *structure* is exactly similar to that of the explanation" (2011: 169). At several junctures, he also explicitly recognizes the speculative nature of his explanation (2011: 13, 166, 168, 170 n. 41, 173 n. 50, 175, 269).

Note that no normative skeptic worth his salt would generously concede it to be an objective normative truth that survival or reproductive success is good. The same applies to Enoch's claim that his explanation would still succeed even if the aim selected for were of no value, provided that some of the things conductive to that aim (such as well-being and feelings of interpersonal trust) are good. Again, no normative skeptic would gratuitously grant that such things are objectively and normatively good. The normative skeptic would argue that the claim "*x* is good" is a normative claim, that as such it is called into question by his evolutionary argument, and that it is therefore something that cannot simply be taken for granted in his rival's counter-argument if this counter-argument it is to be dialectically effective. A moral skeptic would deploy the same line of argument if a third-factor strategy were implemented to defend robust moral realism against evolutionary debunking arguments.

William FitzPatrick (2014) proposes a response to the evolutionary challenge that could be called 'the double-influence argument'. According to this argument, even though the evolutionary moral skeptic presents his explanatory claims about the etiology of our moral beliefs as if they were scientific results, they are not supported by actual science unless it is supplemented with philosophical claims that are question-begging against moral realism. Science only shows that evolution has shaped some of our current moral beliefs to some extent, which leaves open the possibility that other moral beliefs have instead been shaped by systematic reflection that has allowed us to apprehend moral facts, and even that some of the moral beliefs molded by evolution have also been molded by systematic reflection. I think that FitzPatrick is right in this respect, since although moral skeptics like Joyce and Street recognize the hypothetical nature of their arguments, they do not seem to be intellectually humble enough, given the confidence with which they espouse the skeptical conclusions of those arguments. But I think that FitzPatrick himself is guilty of the same intellectual sin, since although his argument may be successful against Joyce's or Street's moral skepticism, it seems to point to Pyrrhonian moral skepticism rather than to non-minimal moral realism. For, as FitzPatrick recognizes, we do not know the extent to which evolution has shaped our current moral beliefs; and we do not know either whether some of our moral beliefs have been shaped only by other factors or whether the beliefs that were molded by evolution have also been molded by other factors such as philosophical, political, or religious reflection. Given our lack of knowledge about the actual extent of the influence

of biological evolution, various kinds of systematic reflection, and experience on our current moral beliefs, it seems that we should suspend judgment about the epistemic status of our moral beliefs. FitzPatrick himself (2014: 247) says that he is not denying the possibility that the debunkers' explanatory claims are correct, in which case it seems, once again, that he should adopt Pyrrhonian moral skepticism. And to FitzPatrick's claim that systematic moral inquiry has given us access to objective facts, one could respond by appealing to the epistemic challenge posed by the version of the argument from disagreement that emphasizes the widespread and entrenched disputes both between first-order moral judgements and between metaethical positions (such as his and Joyce's or Street's), and the difficulty of finding a clear-cut and impartial way of adjudicating such disputes.

I am inclined to think that a moral skepticism of a Pyrrhonian stripe might well represent a more challenging rival to moral realism than other moral skeptical stances, but also a serious rival to these skeptical stances. For the Pyrrhonian moral skeptic recognizes the strength of realist views such as those defended by Enoch and FitzPatrick, but claims that their strength does not appear to be greater than that of skeptical views such as Joyce's or Street's. Besides the strong objections leveled against each of the views in question, those authors explicitly recognize the speculative nature of some of the views they defend and the hypothetical character of the arguments they advance, or acknowledge the possibility that their rivals' views might be correct after all. A Pyrrhonian moral skeptic would wonder how, despite admitting those points, the authors in question can be so confident about the correctness of their views.[18]

3.5. *Moral Projectivism*

The moral skeptic usually recognizes that an important part of his argumentative strategy consists in explaining why human beings naturally believe that there is a moral fact of the matter, and why the great majority of them continue to believe so even after being exposed to what he regards as sound (or at least highly compelling) skeptical arguments. A moral error theorist would put it in these terms: what has led humans, and will continue to lead most of them, to systematically commit the moral error? The argument from evolution discussed in the previous subsection provides one such explanation and even predicts that most people will be dissatisfied with the explanation, given that natural selection has designed the human brain to engage in moral judgment. Another explanation, proposed mainly by Mackie, appeals to the thesis of moral projectivism or objectification.[19] According to this thesis, we project certain sentiments or emotions onto the things, actions, or characters that cause them and are their objects, with the result that we ascribe to those things, actions, or characters certain objective moral features that are intrinsically action-guiding—features that are nonetheless "fictitious." Mackie does not of course mean that we literally project

those sentiments or emotions onto the world, but rather that they cause us to make the above erroneous ascription by making us experience the world as having moral qualities. At one point, he characterizes objectification as "the false belief in the fictitious features" (1980: 72), which can be understood in the sense that the very process of projection consists in coming to erroneously believe in the existence of objective moral features. In talking about "fictitious" moral features, Mackie seems to be already endorsing his ontological moral skepticism. But this makes perfect sense given that moral objectification seems to be proposed as a supplement to his arguments for moral error theory. He complements his error-theoretic position with the objectification thesis in order to offer an explanation of the origin of our pro-morality intuitions that does not imply or presuppose their truth (Joyce 2016e: 187).

An interesting question concerning objectification is that of its relation to an evolutionary account of morality. For instance, a moral error theorist could argue that, while the evolutionary account claims that moral belief was selected for because of its adaptive function or social usefulness, the objectification thesis explains the mental process that gave rise to moral belief. In this case, the process of projection would also be selected for because of its bringing about an evolutionarily advantageous belief in objective moral requirements. Mackie himself remarks that objectification "serves a social function" (1980: 72), and Joyce maintains that it is a plausible hypothesis that natural selection has designed us to project our emotions onto our experience of the world and that the emergence of a projectivist faculty in our ancestors plays a major part in the explanation of the human capacity to make moral judgments (2006: 123–133; cf. 2016b: 10–11).

4. Preview of the Essays

The ten essays in this volume deal with various interrelated issues: error theory, justification skepticism, constructivism, projectivism, veneer theory, inferentialism, disagreement, expressivism, non-naturalism, the Benacerraf challenge, evolutionary debunking arguments, and fictionalism. While some of the essays are sympathetic to moral skepticism, others adopt a critical stance, and still others remain neutral.

Both moral error theorists and moral constructivists maintain that morality is invented or made, not discovered. They differ in that moral error theorists also claim that moral thought is constitutively embroiled in some form of illusion, falsehood, or incoherence inasmuch as it purports to reflect an objective moral reality. In "Projection, Indeterminacy and Moral Skepticism," Hallvard Lillehammer examines the ability of moral constructivism to accommodate two ways in which moral thought has been said by moral error theorists to make erroneous commitments. The first is by believing that, in making moral judgments, one is describing objective moral facts when actually one is merely projecting one's own attitudes onto the

world. The second is by assuming the existence of uniquely correct answers to moral questions that in fact admit of no determinate moral answers. Lillehammer argues that moral constructivism can account for the errors of projection and determinacy that can be attributed to moral thought, because they are only contingent and local phenomena that cannot thereby be taken as universal or necessary traits of moral thought *per se*, contrary to what moral error theorists claim.

Moral error theory can be regarded as a form of *external* skepticism inasmuch as it is a view *about* the moral domain, not a view adopted *within* this domain. It is a metaethical view that seeks to undermine first-order morality and normative ethics 'from the outside' by denying the existence of objective moral truths, facts, or properties that are categorically authoritative. The possibility of external skepticism about morality has recently been contested on the basis of a relaxed conception of truth, fact, and property: claims about the existence of moral truths, facts, or properties are not to be construed as robustly metaphysical, but as moral commitments that are to be assessed according to domain-internal, moral standards. If this were so, then the moral error theorist's denial of such claims would be moral as well, which would render his position incoherent. In "Error Theory, Relaxation and Inferentialism," Christine Tiefensee examines whether a coherent, non-moral form of external skepticism can be constructed, even accepting a relaxed conception of truth, fact, and property. She argues that this is possible provided one adopts an inferentialist construal of moral error theories according to which these theories should refrain from any reference to the alleged falsity of moral judgments and the alleged non-existence of moral truths, and focus instead on claims about the inferential role of moral vocabulary. Tiefensee also assesses the advantages and disadvantages of adopting this new understanding of moral error theories.

Error theory is typically formulated in connection with moral judgments. But there is a more radical version of it that targets normative judgments as a whole, claiming that all such judgments are false. In previous publications, Bart Streumer has argued that we cannot believe such a normative error theory, since it entails that there are no reasons for belief (a reason for a belief being a normative property), and we cannot have a belief and at the same time believe that there is no reason for this belief. But, surprisingly enough, he has also argued that our inability to believe error theory undermines several objections that have been leveled against it and all revisionary alternatives to it, thus making it more likely to be true. In "Why We Really Cannot Believe the Error Theory," Streumer offers a more elaborate version of the argument for the view that we cannot believe error theory and addresses several objections that have been raised to this argument. By doing so, he provides moral error theorists with a new defense against the objection that their theory generalizes to all normative judgments and should therefore be rejected, for the fact that we cannot believe the normative error theory makes this theory more likely to be true.

In Section 2, it was pointed out that there is a weak form of moral realism, commonly dubbed "minimal" or "minimalist," that maintains that moral propositions are truth-apt and some of them are true. In "Are There Substantive Moral Conceptual Truths?" David Copp deals with minimal non-naturalistic moral realism, which he calls "avant-garde non-naturalism." More specifically, he focuses on the type of avant-garde non-naturalism he labels "conceptual non-naturalism," according to which moral non-naturalism does not require that there be non-natural moral properties but only irreducible non-natural moral concepts and substantive conceptual truths involving such concepts. Copp is skeptical about the prospects for such a view, for two reasons. First, the conceptual non-naturalist's theory of concepts faces epistemological and ontological challenges similar to those faced by orthodox moral non-naturalism. Second, so-called moral fixed point propositions are substantive moral propositions but not moral conceptual truths, both because it is implausible that moral error theory is conceptually false and because, as soon as those propositions are qualified to avoid bizarre counter-examples, it becomes clear that they are not conceptual truths.

As noted in Section 2, the notion of objective prescriptivity or bindingness is to all appearances an essential aspect of morality. At the very least, robust moral realists maintain that there exist mind-independent moral facts, properties, or relations that are the source of categorical reasons or inescapable requirements, and one of the reasons moral error theorists target morality is precisely that they take it to posit such queer or mysterious entities. In "The Phenomenology of Moral Authority," Terry Horgan and Mark Timmons agree that categorical authority is an intrinsic phenomenological feature of moral experience and argue that their "cognitivist expressivism" can accommodate such a feature. The key to their argument is the denial that ordinary moral experience is committed to there being objective moral properties and relations that are categorically authoritative, contrary to what both non-naturalistic moral realists and moral error theorists contend. For the claim that there is such a commitment is not revealed as true by direct introspection; nor is it the only possible explanation of those aspects of the phenomenology of categorical moral authority that are reliably revealed by direct introspection. This way, moral phenomenology can exhibit the feature of inescapable authority without there being an ontological error. Still, Horgan and Timmons describe their view as "moderate moral-authority skepticism" inasmuch as it maintains that no moral properties or relations that are categorically authoritative are ever instantiated.

As remarked in Subsection 3.2, the phenomenon of moral disagreement plays a crucial role in certain arguments for either ontological or epistemological moral skepticism. In "Arguments From Moral Disagreement to Moral Skepticism," Richard Joyce critically examines the complex structure of three disagreement-based skeptical arguments with the aim of identifying their difficulties and interrelations. He explores in turn the error-theoretic

version of the argument from moral disagreement that purports to show that there are no (objective) moral facts; the genealogical debunking version that seeks to defeat the justification of moral beliefs; and the version that appeals to the existence of moral disagreement between epistemic peers to undermine the epistemic credentials of their moral judgments. One result of Joyce's analysis is that a common feature of the first two versions of the argument from moral disagreement is that, to succeed in establishing their conclusions, they need to be supplemented with considerations against the plausibility of moral naturalism.

In "Evolutionary Debunking, Realism and Anthropocentric Metasemantics," Mark van Roojen engages with the version of the evolutionary debunking argument according to which it would be a highly unlikely coincidence that our evolutionarily shaped moral beliefs matched the objective moral truths posited by the moral realist. He proposes a new strategy that would make moral realism particularly of a naturalistic stripe immune to that argument. The strategy, modeled on David Hilbert's theory of colors and deploying an externalist metasemantics, consists in distinguishing between the nature of the objective moral properties and our ability to talk about them. Whereas the fact that we are sensitive to moral properties does not make them dependent on us, our epistemic access to them, and hence our ability to talk about them, do depend on our nature. If we had evolved differently, we would not have had epistemic access to, and hence been talking about, the moral properties to which we presently have epistemic access and about which we actually talk. But we would still have had epistemic access to, and been talking about, a different set of properties. Whereas it is a matter of luck that we evolved to grasp, and talk about, the moral properties that we grasp and talk about, it is not a matter of luck that we get things right about them.

The Benacerraf challenge is a well-known objection to Platonism in mathematics. Its proponent argues that, if mathematical entities are, as Platonists claim, mind-independent, causally inert, and existent beyond space and time, then we are led to a skeptical stance according to which it is not possible to explain how it is that we have cognitive access to the mathematical realm or how it is that our mathematical beliefs are reliable. It has been argued that a similar objection could be leveled against those forms of moral realism that fall under what, in Section 2, was called 'robust moral realism'. In "Moral Skepticism and the Benacerraf Challenge," Folke Tersman considers whether, unlike the argument from the best explanation, the argument from disagreement, and the argument from evolution, the moral version of the Benacerraf challenge can undermine moral knowledge without appealing to empirical claims that moral realists deem controversial. His verdict is negative: to successfully counter certain responses to the moral version of the challenge, its proponent needs to have recourse to empirical considerations taken from some of the above arguments.

'Veneer theory' is a term coined by primatologist Frans de Waal to refer to views on which morality is a cultural construction masking people's amoral

biological nature, a nature that is characterized by self-interested (if not selfish) motivation. Veneer theory is a form of moral skepticism inasmuch as it posits rampant hypocrisy while denying the naturalness and prevalence of altruism and non-derivative concerns for justice and fairness. More precisely, it insists with non-cognitivism that moral language is ultimately grounded in command, and joins error theory in maintaining that we are massively ignorant of this fact and mistaken about the ends we are pursuing when issuing or obeying the commands in question. In "Veneer Theory," the most historically oriented of the essays in this volume, Aaron Zimmerman provides a more rigorous definition of veneer theory and questions de Waal's claim that Thomas Henry Huxley was responsible for its adoption by biologists for generations to come. According to Huxley, the in-group solidarity that results from evolutionary group selection is inexorably bound to outgroup hostility. Huxley acknowledged the reality of sympathy, justice, and the other moral capacities de Waal finds among chimpanzees, but Huxley insisted that these moral sentiments are necessarily limited in scope. The 'universal' moral principles trumpeted by eighteenth-century revolutionaries in America and France functioned as a moral veneer, covering the involvement of these same revolutionaries in slavery and imperialism.

It has been claimed that a key difference between ancient and contemporary skepticism is that, unlike the ancient skeptics, contemporary skeptics consider ordinary beliefs to be insulated from skeptical doubt. In the case of metaethics, this issue is related to the following question: what attitude towards ordinary moral thought and discourse should one adopt if one is a moral skeptic? Whereas moral abolitionists claim that one should do away with ordinary moral thought and discourse altogether, moral fictionalists maintain that, given that morality produces practical benefits, one should continue to make moral utterances and have moral thoughts, while at the same time refraining from asserting such utterances and believing such thoughts. Focusing particularly on Mackie's skeptical stance, in "Moral Skepticism, Fictionalism, and Insulation," Diego Machuca considers whether the view that first-order moral beliefs are unaffected by moral skepticism is defensible, whether moral fictionalism is compatible with moral insulation, and whether contemporary moral skeptics are in general committed to there being insulation between first- and second-order levels.[20]

Notes

1 Someone might object that free will skepticism is not a form of skepticism *stricto sensu* inasmuch as most of its advocates deny, rather than doubt, the existence of free will and moral responsibility. My response is that, by doing so, free will skeptics undermine or defeat our commonsense belief that people typically choose and act freely, and are therefore morally responsible for what they do. Likewise, I take atheism to be a form of skepticism inasmuch as it undermines or defeats at least certain common religious beliefs. In the next section, I will deal with a similar objection regarding moral anti-realism.

2 For example, Wright (2013: 1157) criticizes along these lines my inclusion of two essays dealing with moral anti-realism in a volume devoted to the connection between disagreement and skepticism (Machuca 2013).

3 This issue is of course related to the problem of creeping minimalism, on which see e.g. Dreier (2004) and Asay (2013).

4 For defenses of moral naturalism, see e.g. Brink (1989), Railton (2003), and Copp (2007). For defenses of robust moral realism, see e.g. FitzPatrick (2008), Enoch (2011), and Wielenberg (2014).

5 See Mackie (1977: 23–24, 31–35, 40, 42, 59, 73; 1982: 115, 238) and Joyce (2001: 31, 37, 43, 62, 67; 2006: 190–209). See also Garner (1990: 138–139, 145).

6 A natural place to look for confirmation or disconfirmation is research in experimental philosophy and moral psychology. But, unsurprisingly, authors are divided. Some have argued that certain experiments support the claim that children are moral objectivists (Nichols & Folds-Bennett 2003; Wainryb *et al.* 2004), and that moral objectivism is plausibly a default setting on commonsense metaethics, but that at some point in their development a number of people move away from that default view (Nichols 2004). Others have claimed that the folk regard moral beliefs as being almost as objective as factual or scientific beliefs, and considerably more objective than beliefs about social conventions or tastes. But they have also observed that endorsement of moral objectivism varies with such factors as the subject's age, the strength of his opinion about the moral issue, the perceived degree of societal disagreement about it, and whether the moral issue concerns a transgression or an exemplary action (Goodwin & Darley 2008; 2010; 2012; Beebe *et al.* 2015; Beebe & Sackris 2016). And still others have argued that the folk do not have an across-the-board commitment to moral objectivism, since they tend to endorse moral relativism when they consider the views of individuals with radically different cultures or ways of life (Sarkissian *et al.* 2011). Regarding most of the above studies (including the two co-authored by him), Beebe (2015) has raised strong objections both to the way their authors constructed their research materials and to the way they interpreted the results obtained, concluding that "it seems much too early to tell whether and to what degree the folk endorse or reject moral objectivism and what form their objectivism or non-objectivism might take" (28). (What all these authors call "moral objectivism" corresponds to what I call "non-minimal moral realism.") Given the disagreement among experimental philosophers and moral psychologists over the extent of the folk's endorsement of moral objectivism, I prefer to cautiously stick to what I have observed both in my everyday interactions and in the course of teaching an ethics class to a variety of non-philosophy undergraduates for a number of years. Note also that most 'armchair' moral philosophers share the view that the folk are moral objectivists.

7 The earliest proponents of non-cognitivism include Ayer (1936), Stevenson (1937), and Hare (1952). For more recent versions of non-cognitivism, see e.g. Blackburn (1984; 1993), Gibbard (1990; 2003), and Horgan and Timmons (2006). Although the hybrid idea is found in the work of some early non-cognitivists, it has been considerably developed and refined in the last two decades: see e.g. Copp (2001), Ridge (2007), and Boisvert (2008). For recent overviews of moral non-cognitivism, see Schroeder (2010) and van Roojen (2013).

8 I talk about first-order (or basic or substantive) moral judgments to refer to judgments that ascribe a moral property to something or that imply or presuppose the instantiation of a moral property. Thus, judgments such as "There are no objective moral values" or "All substantive moral judgments are epistemically unjustified" do not qualify as first-order moral judgments. Neither do sentences that include moral propositions occurring in unasserted contexts, such as "Paul believes that killing an innocent is morally wrong," "'Killing an innocent

is morally wrong' is a grammatically correct sentence," or "If killing an innocent is morally wrong, then the air force should not bomb the town."

9   For other non-standard versions of moral error theory, see Olson (2014: 6–11, and chs. 2 & 3).

10 Joyce, a prominent moral error theorist, at one point took the label 'moral error theory' to designate both views (2006: 223), but later on abandoned this expansion of the label (2016c: 144 n. 3; 2016d: 161–162).

11 Another argument for moral skepticism could perhaps be constructed on the basis of recent empirical research on the role that disgust and other emotions play in moral judgment. The findings of part of such research challenge the rationalist view, according to which moral judgment is normally the result of a process of conscious reasoning, and support instead the view that it is generally the result of the influence of quick intuitions, moral emotions, and gut feelings. On the latter view, moral reasoning is most of the time an *ex post facto* process in which one seeks arguments that will epistemically justify an already-made judgment with the aim of influencing the intuitions and actions of others. In making the case that the reasons we offer to epistemically justify our moral judgments are usually mere rationalizations, the studies under consideration can be taken to challenge to some extent the epistemic credentials of our moral beliefs, thus supporting an epistemological form of moral skepticism. On the empirical research in question, see e.g. Haidt (2001), Wheatley and Haidt (2005), Haidt and Björklund (2008), and Schnall *et al.* (2008). See Greene *et al.* (2001) for a more nuanced stance on the role that emotions play in moral judgment.

12 For discussion of different versions of the argument from disagreement, see Brink (1984), Tolhurst (1987), Loeb (1998), Tersman (2006), and McGrath (2008).

13 The authors mentioned in the body of the text are sympathetic to the argument from queerness. For a critical assessment of it, see Brink (1984) and Shepski (2008).

14 More precisely, Mackie (1985: 160–161) maintains that the pre-moral tendencies to care for one's offspring and close relatives, to enjoy the company of fellow members of a small group, to exhibit reciprocal altruism, and to display kindly and hostile retribution are to be ascribed to biological evolution. To cultural evolution are to be ascribed "the more specifically moral virtues which presuppose language and other characteristically human capacities and relations" (1985: 161). See also Mackie (1982: 255).

15 More specifically, Street (2006: 119–120) claims that such an influence is indirect: the forces of natural selection have had a great direct influence on our more basic evaluative tendencies, and these tendencies have in turn had a major influence on the content of our evaluative judgments.

16 Street (2008: 208–209) is clearer that (4a) and (4b) are two distinct possible results of (3a).

17 Evolutionary considerations have also played a key role in an argument for skepticism about the moral significance of disgust. Kelly (2011: esp. ch. 5) has argued that this emotion was recruited or co-opted to perform novel functions associated with morality and social interactions, while retaining most of its core structural features that allow it to effectively perform its two primary functions—namely, the avoidance of toxic or poisonous foods and the avoidance of pathogens and parasites. The disgust system provided (additional) motivation to comply with acquired norms and punish those who violated them, and to avoid members of other tribes. Feelings of disgust are therefore irrelevant to the epistemic justification of moral judgments and norms. This form of skepticism is very restricted inasmuch as it is not concerned with the question of whether or not such judgments and norms are epistemically justified.

18 For recent overviews of evolutionary debunking arguments in ethics, see Vavova (2015) and Wielenberg (2016). For an annotated bibliography, see Machuca (Forthcoming). Tersman (Forthcoming) explores how the evolutionary debunking argument and the argument from disagreement can interact.

19 See Mackie (1946: 81–86, 90; 1977: 42–46; 1980: 71–72, 74, 122, 124, 136–138, 147, 149–150; 1982: 239). See also Joyce (2016e; 2016f).

20 I would like to thank Mark van Roojen for his critical comments on a previous version of this chapter.

References

Asay, J. 2013. "Truthmaking, Metaethics, and Creeping Minimalism," *Philosophical Studies* 163: 213–232.

Ayer, A. 1936. *Language, Truth, and Logic*. New York: Dover.

Beebe, J. 2015. "The Empirical Study of Folk Metaethics," *ETYKA* 50: 11–28.

Beebe, J., Qiaoan, R., Wysocki, T., & Endara, M. 2015. "Moral Objectivism in Cross-Cultural Perspective," *Journal of Cognition and Culture* 15: 386–401.

Beebe, J., & Sackris, D. 2016. "Moral Objectivism across the Lifespan," *Philosophical Psychology* 29: 912–929.

Blackburn, S. 1984. *Spreading the Word*. Oxford: Oxford University Press.

———. 1993. *Essays in Quasi-Realism*. Oxford: Oxford University Press.

Boisvert, D. 2008. "Expressive-Assertivism," *Pacific Philosophical Quarterly* 89: 169–203.

Brink, D. 1984. "Moral Realism and the Sceptical Arguments from Disagreement and Queerness," *Australasian Journal of Philosophy* 62: 111–125.

———. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Copp, D. 2001. "Realist-Expressivism: A Neglected Option for Moral Realism," *Social Philosophy and Policy* 18: 1–43.

———. 2007. *Morality in a Natural World*. Cambridge: Cambridge University Press.

Dreier, J. 2004. "Meta-Ethics and the Problem of Creeping Minimalism," *Philosophical Perspectives* 18: 23–44.

Enoch, D. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.

FitzPatrick, W. 2008. "Robust Ethical Realism, Non-Naturalism, and Normativity," *Oxford Studies in Metaethics* 3: 159–205.

———. 2014. "Why There Is No Darwinian Dilemma for Ethical Realism." In M. Bergmann & P. Kain (eds.), *Challenges to Moral and Religious Belief: Disagreement and Evolution*, 237–255. Oxford: Oxford University Press.

Garner, R. 1990. "On the Genuine Queerness of Moral Properties and Facts," *Australasian Journal of Philosophy* 68: 137–146.

Gibbard, A. 1990. *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.

———. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Goodwin, G., & Darley, J. 2008. "The Psychology of Meta-Ethics: Exploring Objectivism," *Cognition* 106: 1339–1366.

———. 2010. "The Perceived Objectivity of Ethical Beliefs: Psychological Findings and Implications for Public Policy," *Review of Philosophy and Psychology* 1: 1–28.

———. 2012. "Why Are Some Moral Beliefs Perceived to Be More Objective Than Others?" *Journal of Experimental Social Psychology* 48: 250–256.

Greene, J., Sommerville, R., Nystrom, L., Darley, J., & Cohen, J. 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293: 2015–2018.

Haidt, J. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108: 814–834.

Haidt, J., & Björklund, F. 2008. "Social Intuitionists Answer Six Questions about Moral Psychology." In W. Sinnott-Armstrong (ed.), *Moral Psychology. Volume 2: The Cognitive Science of Morality: Intuition and Diversity*, 181–217. Cambridge, MA: MIT Press.

Hare, R. 1952. *The Language of Morals*. Oxford: Oxford University Press.

Harman, G. 1977. *The Nature of Morality: An Introduction to Ethics*. Oxford: Oxford University Press.

Horgan, T., & Timmons, M. 1992. "Troubles on Moral Twin Earth: Moral Queerness Revived," *Synthese* 92: 221–260.

———. 2006. "Cognitivist Expressivism." In T. Horgan & M. Timmons (eds.), *Metaethics after Moore*, 255–298. Oxford: Oxford University Press.

Joyce, J., & Kirchin, S. 2010. "Introduction." In R. Joyce & S. Kirchin (eds.), *A World without Values: Essays on John Mackie's Moral Error Theory*, ix–xxiv. Dordrecht: Springer.

Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.

———. 2006. *The Evolution of Morality*. Cambridge, MA: MIT Press.

———. 2016a. *Essays in Moral Skepticism*. Oxford: Oxford University Press.

———. 2016b. "Introduction: Morality: The Evolution of a Myth." In Joyce 2016a, 1–13.

———. 2016c. "Evolution, Truth-Tracking, and Moral Skepticism." In Joyce 2016a, 142–158.

———. 2016d. "Irrealism and the Genealogy of Morals." In Joyce 2016a, 159–174.

———. 2016e. "Patterns of Objectification." In Joyce 2016a, 177–194.

———. 2016f. "Is Moral Projectivism Empirically Tractable?" In Joyce 2016a, 195–218.

Kalderon, M. 2005. *Moral Fictionalism*. Oxford: Oxford University Press.

Kelly, D. 2011. *Yuck! The Nature and Moral Significance of Disgust*. Cambridge, MA: MIT Press.

Loeb, D. 1998. "Moral Realism and the Argument from Disagreement," *Philosophical Studies* 90: 281–303.

Machuca, D. (ed.). 2013. *Disagreement and Skepticism*. New York: Routledge.

———. Forthcoming. "Evolutionary Debunking Arguments in Ethics." In D. Pritchard (ed.), *Oxford Bibliographies in Philosophy*. New York: Oxford University Press.

Mackie, J. L. 1946. "A Refutation of Morals," *Australasian Journal of Psychology and Philosophy* 24: 77–90.

———. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.

———. 1980. *Hume's Moral Theory*. London & New York: Routledge.

———. 1982. *The Miracle of Theism*. Oxford: Oxford University Press.

———. 1985 [1982]. "Co-Operation, Competition, and Moral Philosophy." In his *Persons and Values: Selected Papers*, Volume 2, 152–169. Oxford: Clarendon Press.

McGrath, S. 2008. "Moral Disagreement and Moral Expertise," *Oxford Studies in Metaethics* 3: 87–107.

Nichols, S. 2004. "After Objectivity: An Empirical Study of Moral Judgment," *Philosophical Psychology* 17: 3–26.

Nichols, S., & Folds-Bennett, T. 2003. "Are Children Moral Objectivists? Children's Judgments about Moral and Response-Dependent Properties," *Cognition* 90: B23–B32.

Olson, J. 2014. *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press.

Pollock, J. 1986. *Contemporary Theories of Knowledge*. Towota, NJ: Rowman and Littlefield.

Railton, P. 2003. *Facts, Values, and Norms*. Cambridge: Cambridge University Press.

Ridge, M. 2007. "Ecumenical Expressivism: The Best of Both Worlds," *Oxford Studies in Metaethics* 2: 51–76.

Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. 2011. "Folk Moral Relativism," *Mind & Language* 26: 482–505.

Schnall, S., Haidt, J., Clore, G., & Jordan, A. 2008. "Disgust as Embodied Moral Judgment," *Personality and Social Psychology Bulletin* 34: 1096–1109.

Schroeder, M. 2010. *Noncognitivism in Ethics*. London: Routledge.

Shepski, L. 2008. "The Vanishing Argument from Queerness," *Australasian Journal of Philosophy* 86: 371–387.

Sinnott-Armstrong, W. 2006. *Moral Skepticisms*. New York: Oxford University Press.

Stevenson, C. 1937. "The Emotive Meaning of Ethical Terms," *Mind* 46: 14–31.

Street, S. 2006. "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127: 109–166.

———. 2008. "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying about," *Philosophical Issues* 18: 207–228.

Tersman, F. 2006. *Moral Disagreement*. Cambridge: Cambridge University Press.

———. Forthcoming. "Debunking and Disagreement," *Noûs*. DOI:10.1111/nous.12135.

Tolhurst, W. 1987. "The Argument from Moral Disagreement," *Ethics* 97: 610–621.

van Roojen, M. 2013. "Moral Cognitivism vs. Non-Cognitivism." In E. Zalga (ed.), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/archives/fall2015/entries/moral-cognitivism/.

Vavova, K. 2015. "Evolutionary Debunking of Moral Realism," *Philosophy Compass* 10: 104–116.

Wainryb, C., Shaw, L., Langley, M., Cottam, K., & Lewis, R. 2004. "Children's Thinking about Diversity of Belief in the Early School Years: Judgments of Relativism, Tolerance, and Disagreeing Persons," *Child Development* 75: 687–703.

Wheatley, T., & Haidt, J. 2005. "Hypnotically Induced Disgust Makes Moral Judgments More Severe," *Psychological Science* 16: 780–784.

Wielenberg, E. 2014. *Robust Ethics: The Metaphysics and Epistemology of Godless Normative Realism*. New York: Oxford University Press.

———. 2016. "Ethics and Evolutionary Theory," *Analysis* 76: 502–515.

Wright, S. 2013. "Review of Machuca 2013," *Mind* 122: 1157–1160.

2 Projection, Indeterminacy and Moral Skepticism

Hallvard Lillehammer

1. Moral Constructivism and Moral Error Theory

According to one way of understanding moral skepticism, to be a skeptic about morality is to hold that all moral claims are either false, incoherent, or something else misleadingly expressed. This tripartite disjunctive formulation is due to Bernard Williams, who made this claim about attributions to agents of reasons for action said to obtain regardless of the contents of their desires, and hence about any moral reasons so said to obtain (Williams 1981). Williams's formulation (although not his entire view) broadly captures the central claims of other recent proponents of moral skepticism, such as John Mackie (1977), Richard Joyce (2001), and Jonas Olson (2014)—also commonly known as 'moral error theorists'. In what follows, I use the terms 'moral skepticism' and 'moral error theory' interchangeably.

According to moral error theory, morality is something invented, constructed or made, but mistakenly presents itself to us as if it were an independent object of discovery. According to moral constructivism, morality is something invented, constructed or made.[1] Thus understood, error theory and constructivism are close philosophical relatives (Lillehammer 2011). Both types of view take morality to be a construction. But error theorists go further than constructivists in claiming that moral thought necessarily, or 'constitutively', presents itself as something different, or 'more' than that, namely, a cognitive reflection of a moral reality that exists independently of our contingently evolved and historically emergent beliefs, desires, institutions or social practices.

The fact that constructivism stops short of claiming that morality presents itself as something it is not might be thought to give it a significant explanatory advantage over error theory insofar as the constructivist thereby avoids lumbering moral thought as a whole with a set of allegedly dubious commitments (see, e.g., the papers collected in Lenman & Shemmer 2012). As a starting point for discussion, this might indeed be a reasonable assumption. Yet its avoidance of attributing these allegedly dubious commitments to moral thought as a whole is only a genuine explanatory advantage of constructivism if at least two further conditions are met, namely,

that (*i*) the constitutive presence in moral thought of these allegedly dubious commitments is not supported by the evidence of how moral thought actually works, and that (*ii*) a constructivist account of moral thought is not affected by other explanatory disadvantages when compared to its error-theoretic competitors. In what follows, I have very little to say about (*i*).[2] Instead, I focus on (*ii*). In particular, I focus on the hypothesis that the ability of error theories to diagnose a range of genuinely dubious commitments embodied in moral thought is a comparatively neglected argument in their favor, as compared to constructivist accounts that have historically had relatively little to say about this aspect of moral thought. The cause of this explanatory gap in standard constructivist accounts is a topic I do not go into great detail about here, although it is likely to be at least partly a consequence of the fact that historical proponents of constructivism have understandably been more concerned to vindicate allegedly dubious aspects of moral thought than to debunk them. (I return to this point very briefly below.)

In what follows, I argue that, when suitably understood, constructivism is both compatible with, and in certain cases explanatory of, some of the allegedly dubious commitments to which arguments for moral skepticism appeal.[3] What distinguishes plausible constructivist accounts from paradigmatic versions of error theory is a diagnosis of these dubious commitments as contingent or local, and as existing against a background of moral claims to which the error-theoretic hypothesis is not taken to apply (or is not taken to apply for the same reasons). In this essay, I focus on two particular allegations that are sometimes associated with moral skepticism.[4] The first is the suspicion that in making moral claims, we are merely 'projecting' our attitudes onto the world, without there being any moral fact of the matter to which these attitudes correctly respond. I refer to this as the 'projection challenge'. The second is the suspicion that in arguing for and against moral views, we are merely attempting to influence each other to give similar, overlapping or identical answers to questions which in themselves have no determinate moral answer. I refer to this as the 'indeterminacy challenge'. While taking it for granted that some projection and indeterminacy errors actually do occur in moral thought, I describe two ways in which the existence of a restricted error-theoretical hypothesis can be incorporated by a constructivist account. On such a view, projection or indeterminacy can (and often do) obtain in a manner affecting the soundness of moral thought, but do so as contingent and local features of moral thought—not universal or necessary ones. Moreover, insofar as either projection or indeterminacy obtain in a manner affecting the soundness of moral thought, the challenges they present will often be substantially moral ones, and will therefore be possible to address from within moral thought itself. Taken together, these conclusions imply a reduction in the number of questions in metaethics that threaten to falsify or otherwise debunk the commitments embodied in moral thought as a whole, yet stop short of supporting the more radical view of those who

claim that all questions in metaethics are reducible to questions of 'first-order morality' without remainder (cf. Kramer 2010; Dworkin 2011).

To avoid a basic misunderstanding at the outset, I should stress that being able to show how a constructivist account can accommodate the projection and indeterminacy challenges (and how the resulting account offers a diagnosis of what on constructivist terms are the excessively general implications drawn from these challenges by the moral skeptic) does not constitute a refutation of moral skepticism on its own terms. On the other hand, it might conceivably reduce the temptation (if such a temptation were felt) to think in those terms in the first place. If refuting the moral skeptic on his or her own terms were possible (and I am not sure that it is), that would be the topic of a different paper. I shall, however, return to the issue briefly in Section 3 below.

2.   Projection and Indeterminacy

2.1.   *Projection*

Sometimes we seem to attribute to aspects of our social world features of our own responses to that world, thereby mistaking the effect for the cause. This is a widely recognized epistemic defect, and one that sometimes goes under the name of 'projection' (see, e.g., Kail 2007). Projection so understood can also be a moral defect, as exemplified by a vindictive person thirsty for revenge who falsely attributes to the object of their vindictiveness a hostile, aggressive or otherwise unsavory attitude to others. To the extent that we project our own attitudes onto the social world in this distinctive way, we are apt to act less than well, for example by supporting retributive action against people who are completely blameless in the circumstances and who might actually be the object of the same hostile, aggressive or otherwise unsavory attitudes themselves. Projective error thus understood does not require its subject to be suffering from ill-intention or bad character. To take a slightly different example, it is possible for me to get so absorbed in my enthusiasm for a certain charitable plan or policy that I mistakenly come to think my favoring of it is in line with the general enthusiasm that plan or policy elicits in others, when in fact it elicits none. Thus, many people are familiar with the shock of the person who is told by her interlocutors that what seems so obviously agreeable to her is, as a matter of fact, completely unacceptable to everyone else, and where the difference in question comes down to the different affective construal by different people of the case in hand (e.g., where one person sees an opportunity, another person senses danger, and so on).

At least some constructivist accounts of moral thought agree that all moral judgments involve some element of 'projection', if in somewhat different ways.[5] Thus, it has been argued that all sound monadic predications of moral qualities to actions or states of affairs (such as goodness or

badness) are best interpreted so as to include a hidden, or implicit, relational element connecting specific features of the social world with a set of attitudinal responses to that world in standard or otherwise suitable circumstances, such as a tendency to like or dislike the thought of their coming to exist (see, e.g., McDowell 1998; Wiggins 1998; Gert 2012). Yet even if all moral judgments involve a comparatively innocuous 'projection' of attitude in this perfectly general way, there will remain a subset of moral judgments (such as those described in the previous paragraph) that involve the 'projection' of attitudes in a distinctively more problematic way. Yet in favorable circumstances, moral judgments that do involve a 'projection' of attitudes in this distinctively more problematic way can in principle be reflectively updated or checked against the facts, such as facts about the causal history and function of those judgments (e.g., when I realize that the person actually exhibiting the hostile, aggressive or otherwise unsavory attitude is not my victim, but myself). Moral judgments that either survive or emerge from such a reflective process unscathed are ones a constructivist may classify as 'reflectively robust', as opposed to ones that would not (where neither set is thought to be empty).[6] The fact that all moral judgments involve the comparatively innocuous 'projection' of attitudes in the perfectly general way does not entail that all moral judgments involve a form of 'projection' in the distinctively more problematic way. This is not to say that a sufficiently informed explanation of the 'projective' aspects of moral thought could not in principle support a general form of moral skepticism. Clearly it could, if all moral judgments could be shown to be 'projective' in the distinctively more problematic way described in the previous paragraph. Yet as it happens, at least some moral judgments do seem to be reflectively robust in the sense that they either do *not* involve the distinctively more problematic form of 'projection' described in the previous paragraph, or that a commitment to those judgments *would* survive our discovery of their original dependence on this distinctively more problematic form of 'projection'. (Having discovered that your initial judgment was a product of unjustified prejudice, for example, you might in principle uncover alternative reasons for sticking to that judgment.) It follows that it is more than merely wishful thinking to believe that even though some 'projective' errors genuinely do obtain in moral thought, they do not (and so do not necessarily) affect all moral judgments equally; even if all moral judgments do involve some element of 'projection' in the comparatively innocuous and perfectly general way admitted to obtain by at least some moral constructivists.[7]

2.2.   *Indeterminacy*

The above description of the reflective robustness of moral claims runs straight into an obvious challenge. What seems reflectively robust for me in some situation may not be what seems reflectively robust for me in another situation, or what seems reflectively robust for you in this or in any other

situation. As we all know, people holding what seems to them to be reflectively robust moral beliefs will sometimes disagree, even in the knowledge that the people they disagree with claim to be in exactly the same epistemically strong position as they do. What if each set of conflicting moral beliefs is reflectively robust to the same (maximal) degree? What if there is nothing to decide between them? What if the question is too vague, the values involved not possible to compare, or the answers to the moral questions at issue indeterminate? In that case, we are no further, assuming (as the moral skeptic might surmise) that in making moral judgments we necessarily take for granted that moral questions do have determinately correct answers that are the same for everyone in all circumstances.

Sometimes people do insist that their answer to a moral question is the uniquely correct one even if there are, in fact, other ways of answering the question that are equally sensible, or just as good. In each case, they are likely to be exhibiting an epistemic defect, even if they are lucky enough to have hit upon one of the acceptable answers. The epistemic defect in question is one of dogmatism, which can result from confusing the reasonable desire to find an acceptable answer to a pressing question with the unreasonable insistence that there can be only one. This kind of dogmatism can also be a moral defect, as when someone adopts an intolerant attitude towards other people whose approach to an issue is perfectly reasonable, but significantly different from their own (as may happen when someone falls for the temptation to place victory in argument above the virtues of fairness or truthfulness). To take one actual example among many, some of us are familiar with institutional norms and regulations that differ subtly from one institution to another (e.g., between different clubs, societies, departments or corporations of the same kind), where each set of norms and regulations actually serves the ends of the institutions that adopt them as well as could reasonably be hoped for, in spite of being extensionally different from each other (e.g., in terms of how votes are accumulated when electing officers in one of the many selection practices that go under the label 'democratic'). In each of the institutions in question, there may have been long (and sometimes tedious) internal discussions between members about which set of norms and regulations is the correct one to adopt, discussions that may have taken place in more or less complete isolation from exactly parallel discussions in the institution next door. In each case, the debates may have been intense, and the disagreements deep and trenchant. At the same time, it might be the case that for each set of mutually incompatible norms and regulations considered in these discussions, at least one of the relevant institutions has actually gone on to adopt it, and have done so successfully. Hence, by all accounts, and in spite of how things may have looked to the local disputants at the time, there could be more than one set of norms and regulations that is perfectly acceptable for the kind of institution in question—norms and regulations that would reflectively stand the test of time, and that are capable of justly and effectively serving the institutions that endorse them.

This in spite of the fact that each set of norms and regulations will conflict with all the others on at least one issue. Once the virtues of epistemic modesty (e.g., always bearing in mind the possibility that further discussion or new evidence may decide the issue) and moral integrity (e.g., bearing in mind that some forms of pluralism are a symptom of intellectual laziness or complicity by omission) have been taken into account, to insist that only one of these sets of norms or regulations could in fact (or 'really') be correct for institutions of this kind can be a symptom of an insufficient grasp of the scope of acceptable difference in morally legitimate and worthwhile social practices, where the question of whether to accept as adequate more than one solution to a problem is sometimes itself a substantially moral one.

According to moral error theory, all substantial moral judgments are either false, incoherent, or something else misleadingly expressed. Hence, there are no correct answers to moral questions, never mind a unique and determinate answer in every (or, indeed, any) case. If there is no right answer, then all moral questions are objectively indeterminate at best, any solution to a moral problem being as good as any other, at least as far as its distinctively moral merits are concerned. When it comes to getting moral things right (as opposed to getting other things right, such as being able to survive, making a living, or silencing one's opponents, etc.), there is literally no issue of correctness to discuss.

There is a good case to be made for the claim that the answer to at least some moral questions is indeterminate (cf. Putnam 2004; Albertzart 2014: ch. 4; Constantinescu 2014). On that issue, the moral skeptic is probably right. These questions include ones that admit of more than one perfectly reasonable solution, possibly like the institutional examples I sketched in the previous paragraph. There is also some evidence that at least in some cases people do make the false assumption either that some specific answer to a moral question is uniquely and determinately correct, or that there must be such an answer, even if there is none. Yet as we attain some degree of moral maturity, we normally become able (perhaps after a bit of agonistic soul-searching or criticism from others) to grasp that sometimes a question that previously seemed to us to allow for one, and only one, acceptable answer does, in fact, have a plurality of acceptable answers. One way for there to be a plurality of acceptable answers to a moral question is for two or more answers to be most reasonable together, or disjunctively. Thus, to the question whether we should choose A or B, the answer could be: choose either (e.g., "Take a bath in the sea or a cold shower"). More importantly for present purposes, many plausible cases of apparent moral indeterminacy are located against a background of generally accepted determinacy. Thus, to the question whether we should choose A or B, the answer may be: choose either, but definitely not C or D (e.g., "Either give her an aspirin or a paracetamol, but don't give her all the pills in either bottle all at once"). Thinking back to the institutional case described in the previous paragraph, there is normally a limited range of possible norms and regulations within

which serious participants could reasonably favor one set over another. Beyond that range, there is a set of conceivable norms and regulations the lack of coherence of which, the self-defeating nature of which, or the sheer perversity of which will make morally serious participants either rule them out from the start, or (so one might wish) never think of them in the first place. To accept that some moral questions have no unique, determinate and non-disjunctive answer does not therefore imply that we should worry about losing our moral bearings altogether, even if some claims to unique and determinate correctness are either silly or dogmatic, or both.[8] Finally, the fact that people do sometimes falsely insist on there being a unique, determinate and non-disjunctively correct answer to some moral questions does not support the error-theoretic hypothesis that moral thought as such is defective in virtue of its false insistence on there being unique, determinate and non-disjunctively correct answers to all moral questions. So long as only a single moral question has a unique, determinate and non-disjunctively correct answer, the unrestricted error-theoretic hypothesis is false. Moral thought can survive the case-by-case discovery of local indeterminacy with respect to particular sets of moral claims against the background of moral claims, the (in-)correctness of which is perfectly determinate. In this sense, the discovery of moral indeterminacy is potentially 'conservative' with respect to the status of the moral judgments against the background of which the indeterminacy in question is located.

The local, or even widespread, existence of indeterminacy in moral thought is readily explicable on any constructivist account on which moralities (like social institutions, such as clubs or societies) are understood as individual and/or collective works in progress or development, and to this extent potentially undecided, open-ended or disjunctively permissive with respect to at least some moral claims that either have been, or could be, coherently formulated.[9] Such constructivist accounts can readily account for even the widespread existence of mistaken commitments to unique, determinate and non-disjunctively correct answers to moral questions while stopping short of the skeptical claim that all moral thought is thus mistakenly committed with respect to every moral judgment that could possibly be made. The range of constructivist accounts able to accommodate this phenomenon may fail to include certain views according to which the answer to all moral questions is said to be fixed as a matter of necessity by features constitutive of human (or even rational) agency. Yet it is hard to think of any form of constructivism, even of the strictest Kantian variety, that entails this radical conclusion (see, e.g., Korsgaard 2009). Either way, the requisite degree of indeterminacy in moral thought is (*i*) consistent with any form of constructivism according to which the correctness of moral judgment is taken to be a function of moral thought and practice, or its historical development; and (*ii*) explained by any form of constructivism according to which moral thought and practice, or its historical development, is partially constitutive of the correctness of moral judgments.

2.3. Projection and Indeterminacy

The false projection of idiosyncratic responses onto a non-obliging social reality and the mistaken insistence on there being unique, determinate and non-disjunctively correct answers to moral questions are logically independent ways for moral judgments to be in error. Yet there are potentially illuminating ways in which these two errors could be combined to produce a distinctive source of error in moral thought. One way in which some moral judgments could involve a distinctive error so understood is as follows. Suppose my reaction to a morally significant situation (such as a proposed change in my working patterns) is fear (e.g., because I tend to be averse to changes proposed by people other than myself). I judge the proposed change to be a risky and dangerous threat to what is agreed by all to be a collegial and mutually supportive working environment, and strongly object to it on those grounds. However, most of my colleagues fail to share this reaction, not having experienced the same levels of fear or surprise as I. In fact, most of them see the proposed change as a positive step, having as a predictable consequence the saving of person hours and a prospect for all employees to spend more time with their families. A heated dispute ensues, in my own case characterized by the following two key facts. First, I am adamant that the proposed change is bad, in virtue of being a serious threat to the existing collegial and mutually supportive office culture (at least as I imagine it to be). Second, I am adamant that there is no other acceptable solution than the *status quo*. There can be one, and only one, reasonable solution to this conflict, and that is the one I have proposed. (The judgments of my colleagues may or may not exhibit the same dogmatic tendencies.)

This imaginary (but not entirely unrealistic) example shows two sources of moral error working together, with the first error (projection) playing an explanatory role in generating the second (determinacy). In short, my mistaken insistence that there is a unique, determinate and non-disjunctively correct answer to the question at hand is in this case explained by my mistaken insistence that the only salient alternative to my own view is dangerous, where the latter judgment is explained by my unreasonable fear. In this way, an 'error of projection' and an 'error of determinacy' can (and sometimes probably do) work together to produce a more complex and distinctive error in moral thought in the same situation. A constructivist account of moral thought can accommodate this kind of complex error. Yet on a plausible constructivist account, errors of this kind will be interpreted as the contingent, local and in principle correctable errors which in at least some cases they hopefully are.

2.4. Comparisons and Implications

What I have called the 'error of projection' essentially involves the mistaken attribution to the social world of a feature that is actually located in the mind of the person making the relevant moral claim, as when my fear causes me to regard some harmless aspect of the world as dangerous.[10] The error of

determinacy does not necessarily involve a mistaken attribution of this kind. Yet the error of determinacy could also involve a mistaken attribution to the world of a feature that occurs only in the mind of the person making a moral claim, such as would be the case where my belief that some moral question must have a unique, determinate and non-disjunctively correct answer is explained by the fact that the possibility of things being as I think they are is one I find too painful to contemplate. While the latter case may not be helpfully described as a case of 'projection', it certainly belongs to a similar family of epistemic (and sometimes moral) defects, in the sense that the way I judge the world to be is excessively colored by how I feel. Errors of determinacy essentially consist in overestimating the size of the domain of questions that have a unique, determinate and non-disjunctively correct answer. Thus understood, they are errors of over-generalization. Errors of projection are not characterized by this kind of over-generalization, as opposed to falsely locating one set of features in the wrong place (e.g., 'the world' as opposed to 'the mind'), and potentially mischaracterizing those features in so doing (e.g., as 'danger' as opposed to 'fear'). In these respects, errors of projection and determinacy are importantly different, both with respect to what they consist in, and with respect to the kind of diagnosis and correction they are likely to be susceptible to during the course of moral reflection and development.

Errors of projection and determinacy also differ in the way their emergence or persistence in moral thought can be given a plausible functional explanation or rationale. It is comparatively easy to see why we should not be surprised to find errors of determinacy in various areas of moral (and other) thought, insofar as a commitment to argue as though the truth is "single" (cf. Blackburn 1985) can function to overcome epistemic complacency; to generate such agreement as is possible; to discover non-obvious, but reasonable alternatives; to increase the sophistication and overall coherence of our moral beliefs; and to ensure that we leave no stone unturned in matters of great importance, and so on. To that extent, it might be tempting to describe the error of determinacy as manifesting a 'virtue' that involves a false belief at its core. A tendency to mistake our groundless fear for genuine danger, on the other hand, is not so obviously a recipe for sharpening our moral/epistemic tools, nor for getting better at discovering such genuine danger as our social world clearly contains. To this extent, it may be harder to think of the error of projection as manifesting an epistemologically or morally admirable outlook. Yet in some situations, even a tendency to falsely 'project' an idiosyncratic attitude onto the social world could arguably play a morally significant social role, for example if that social world is widely and accurately characterized by my merited distrust, and where the false 'projection' of fear, for example, comes to serve a protective function. Furthermore, by projecting my own idiosyncratic responses onto the social world, I may succeed in 'framing' the way that other people approach it so as to encourage responses that serve my ends.[11] To this extent, errors of projection can also be given a morally significant rationale. It follows

that unless they are (jointly or separately) replaceable without great cost by moralized responses that play a similar role without the associated error, both errors of projection and errors of indeterminacy can in principle play the role of self-reinforcing (but contingent and local) 'fictions', in something like the way that some error theorists have claimed that all moral claims do (cf. Joyce 2001; Lillehammer 2004).

From a constructivist perspective, the functional explanation or rationale proposed for errors of projection and determinacy will differ significantly from the parallel explanation or rationale proposed by standard forms of error theory. Most significantly, standard forms of error theory will characteristically attempt to give a functional explanation or rationale for the alleged error in question that is articulated reductively, and so entirely in non-moral (and possibly non-normative) terms. From the perspective of constructivism, the situation is different. Because constructivism allows that a wide range of moral claims are actually true, valid or otherwise correct as they stand, it will be natural (and perfectly consistent) for a constructivist account of moral thought to give a functional explanation or rationale for local errors in moral thought that is either explicitly or implicitly moral to some extent. In other words, given that some moral claims are actually reflectively robust, it will be natural (and perfectly consistent) for a constructivist to say that some errors of projection or indeterminacy have a substantially moral function or rationale, and to treat that as a genuine explanation to be evaluated on its own (partly moral) terms. Any analogously moralized explanations offered by a moral skeptic, on the other hand, must be seen for what they are, namely, as at best second-rate approximations to explanations, useful explanatory illusions, or some other kind of explanatory 'Fools' Gold'.

3.   Some Preliminary 'Results'

There is such a thing as 'errors of projection' in moral thought. One error of projection consists in misattributing aspects of one's own personal attitudes to features of the social world, for example by overestimating the extent to which those attitudes are shared by others. This kind of error of projection can be corrected for by investigating the facts in question, including relevant facts about the causes and function of one's own attitudes. The existence of this kind of error of projection does not support a general form of moral skepticism. In particular, it does not support the skeptical conclusion that all moral claims involve an error of projection that embroils them, of necessity, in falsehood, incoherence or misunderstanding.

There is also such a thing as an 'error of determinacy' in moral thought. One kind of error of determinacy consists in overestimating the extent to which moral questions have unique, determinate and non-disjunctively correct answers, sometimes by insisting that 'moral truth' is single in the face of stronger evidence for either pluralism or indeterminacy. This kind

of error of determinacy can also be corrected for by investigating the facts in question, including the relevant facts about the causes and function of the insistence on perfect determinacy. The existence of this kind of error of determinacy does not on its own support a general form of moral skepticism either. In particular, it does not support the skeptical conclusion that all moral judgments are falsely committed to the view that all moral questions have unique, determinate and non-disjunctively correct answers.

Both of these preliminary conclusions are consistent with the claim that some moral judgments do embody genuinely dubious commitments. In some cases, these will be dubious 'non-moral' commitments (e.g., that the non-human world is literally 'hostile' or 'friendly' to our personal desires or concerns, as opposed to either 'dangerous' or 'beneficial'). In other cases, they will be dubious 'moral' commitments (e.g., that it is always right to insist that there can be only one acceptable solution to a coordination problem, when in fact there can be more than one). There are at least three ways to go wrong in giving an account of the place of such dubious commitments in moral thought. The first is to treat local and contingent features of some moral judgments as if they were necessary features of all moral judgments (cf. Williams 1985; Lillehammer 2013). The second is to treat substantially moral questions as if they were 'non-moral' questions about the nature of morality (cf. Blackburn 1998; Kramer 2010). The third is to overplay the significance of the fact that some moral judgments commit us to a 'disorderly', or otherwise insufficiently sparse, 'ontology' (cf. Putnam 2004; Price 2011). Although visible at the margins, the third of these ways of going wrong has been much less prominent in the preceding pages than the first two. This is not to deny that it is of genuine interest (cf. Lillehammer 2013).

It might be objected that the descriptions I have given in the two previous sections of local and contingent errors in moral thought are poorly targeted as far as traditional discussions of moral skepticism are concerned. These discussions, far from making invalid inferences from local and contingent aspects of some moral claims, explicitly address what are claimed to be universal or necessary features of such claims, such as an allegedly false commitment to mind independence, or an incoherent postulation of irreducible normative properties (Mackie 1977; Joyce 2001; Olson 2014). With respect to these questions, what I have described in the previous two sections as errors of projection and determinacy are beside the point.

There is inevitably some truth in this complaint. But then my aim in this essay has not been to produce an argument against moral skepticism on its own terms (which we may or may not accept as a sensible starting point), but instead to show that one feature of moral thought that might seem to be better accounted for by moral skepticism can in fact be accounted for also on a moral constructivist account, *up to a point*. (The qualification 'up to a point' is crucial because from a constructivist perspective, there is no phenomenon to be explained beyond the relevant point, and so no rationale for making the attempt.) Thus, I have tried to be very clear throughout that

my main focus would not be every allegedly universal and necessary feature of moral claims that has historically been targeted in existing arguments for moral skepticism, but rather a more restricted domain of contingent, local and generally correctable errors that can be agreed to exist by moral skeptics and their critics alike. The reason for choosing this focus is precisely that in discussing these contingent, local and generally correctable errors, it is possible to show not only that constructivist accounts of moral thought are able to capture some of the genuine insights about projection and indeterminacy alluded to in standard arguments for moral skepticism, but also to raise the possibility that in some cases at least, those insights are more plausibly interpretable as capturing contingent, local and correctable errors, as opposed to the necessary, universal and non-eliminable errors that moral skeptics have claimed to find in moral thought as a whole. Finally, the errors of projection and determinacy I have described in the previous two sections are recognizably real and empirically observable errors in moral thought, as opposed to at least some of the errors claimed to exist by recent moral theorists—errors which have tended to be visible only on the analytical metaphysician's balance sheet, and the alleged implications of which have therefore tended to be restricted to the ordering of the columns therein (cf. Blackburn 1998). The errors of projection and indeterminacy I have described in the previous two sections should therefore be of interest also to those who are concerned about the threat of so-called 'debunking arguments' in ethics more widely, regardless of the comparative merits of constructivism and error theory as comprehensive theories in metaethics.

4.   In Search of the Moral Skeptic

According to one form of ancient skepticism, there is no more reason to believe any proposition than there is to believe its contradiction. We should therefore be agnostic with respect to the truth value of all propositions, moral or non-moral, at the cost of irrationality. Regardless of the well-known issue about whether it is internally coherent (i.e. should we be agnostic also with respect to the skeptical proposition?), this form of skepticism, widely known as 'Pyrrhonian', has one particularly interesting feature. The Pyrrhonian skeptic does not attempt to stand outside of ordinary discourse in order to judge it as epistemically defective with respect to some externally validated standard. Instead, the skeptical conclusion is derived from inside ordinary discourse by considering each claim and counterclaim on a case-by-case basis, and finding that neither can be more or less securely established than the other. In this sense, Pyrrhonian skepticism is a paradigmatic form of 'internal' skepticism. It is therefore not refutable by the philosophical strategy of pointing to the alleged incoherence, absurdity or impracticality of taking up a philosophical perspective (or an 'Archimedean point') outside of human thought and practice altogether (cf. Williams 1985).

Pyrrhonian skepticism has a metaphysical analogue with which it should not be confused (but with which it is arguably interchangeable in practice). On this view, there is no proposition that is more true, valid, or otherwise correct than its contradiction. We should therefore stop short of actually believing any proposition, moral or non-moral, at the cost of illusion. Unlike its epistemic analogue, this form of skepticism unquestionably deserves the title of an 'error theory' about said propositions, the implication being that to believe a proposition is to believe either a falsehood, a contradiction, or something else that falls short of the 'aim' of truth. Like its epistemic analogue, this form of skepticism does not attempt to stand outside of ordinary human discourse in order to judge it metaphysically defective with respect to some externally validated ontological standard. Instead, the skeptical conclusion is said to be derivable from the inside of ordinary human discourse by considering each claim and counterclaim on a case-by-case basis, and finding that neither can be true (e.g., because they both entail a contradiction). In this sense, the metaphysical analogue of Pyrrhonian skepticism is also a paradigmatic form of 'internal' skepticism. It is therefore not refutable by pointing to either the incoherence, absurdity or impracticality of taking up a philosophical perspective (or an 'Archimedean point') outside of human thought and practice altogether.

A constructivist account of moral thought is compatible with 'internal' forms of moral skepticism, of both the epistemic and the metaphysical variety. Yet if the arguments I have given in this essay are plausible, we can make sense of this kind of 'internal' moral skepticism as contingent and local. A constructivist account of moral thought is also compatible with the claim that it is possible to ask 'external' questions about moral thought, such as how moral claims can be 'located' with respect to other claims to which we attribute justification, truth or facticity. Yet if the claims I have made in this essay are at all plausible, any such 'external' location exercise is likely to be of limited significance for what we should say about our entitlement to claims about moral justification, truth or facticity, as opposed to what we should say about how such claims are most plausibly interpreted (cf. Blackburn 1998). As far as I can see, this is all as it should be. The Pyrrhonian hypothesis is one that ought to be rejected—on epistemological, on metaphysical, and on substantially moral grounds.[12]

Notes

1 I intend this characterization of constructivism and error theory to be consistent with the claim that significant aspects of the human moral sensibility are a product of evolutionary and other non-intentional causes (see e.g. Lillehammer 2003). I mainly ignore this complication in what follows.
2 For a discussion of (*i*) in the context of current controversies, see e.g. Finlay (2014).
3 I do not argue, nor would I want to claim, that constructivism is the only form of metaethical theory that can offer such explanations. Any minimally plausible metaethical theory should have the resources to offer such explanations. I take

at least some versions of moral realism to be at least minimally plausible in the relevant sense. For example, some of the explanatory power of constructivist accounts of indeterminacy and permissiveness (discussed below) might be equally shared by realist views that accommodate indeterminacy and permissiveness by appealing to the vagueness inherent in (some) moral predicates.

4 Comprehensively executing the strategy employed in this chapter would require addressing every allegedly dubious commitment that either has been, or plausibly could be, targeted by moral error theorists, including such familiar suspects as the idea of a categorically binding reason for action, an intrinsically 'magnetic' normative fact, and so on. Addressing every allegedly dubious claim that has been targeted by moral error theorists is an unrealistic ambition for a single essay, so I do not attempt it here.

5 The aim of the discussion that follows in the main text is to sketch how some constructivist accounts can allow for certain errors of projection, not to establish the independent plausibility of any form of constructivism. I therefore allow myself to skirt over a number of details of the different ways in which a constructivist account of moral thought could plausibly be developed, except where doing so is clearly in danger of making my conclusions look more general than they are intended to be. For further discussion of different kinds of constructivism, see e.g. Lillehammer (2011) and the essays in Lenman and Shemmer (2012).

6 There are interesting complications here because some forms of 'projection' in the distinctively more problematic way can play a crucial role in the promotion of ends we are unwilling to renounce, even after suitably informed reflection. I briefly return to this issue in my discussion of self-reinforcing 'fictions' in Section 2.4 below.

7 The line of thought I have sketched in the main text would not go through for meta-ethical accounts that postulate correctness conditions for moral judgments that are said to exist in complete independence of moral thought—or at least it would not go through for such accounts without the addition of further (controversial) premises. It is partly the fact that the form of constructivism in question construes the correctness of moral judgments as some function of moral thought itself that makes it plausible to insist on the crucial distinction between the two senses of 'projection' in the way I have done in the main text. For further discussion of the implications of this point, see e.g. Lillehammer (2003) and Street (2006).

8 To say that a unique and determinate answer is 'non-disjunctive' is not to deny the obvious truths that there is always more than one way to carry out an action, and that any proposition can trivially be turned into a disjunction without change in truth value. The implicit assumption is that we can identify a subset of disjunctions as informative or interesting for the purposes at hand. To this extent, the notion of a 'non-disjunctive' answer is correspondingly theory- or interest-relative.

9 The basic sketch of constructivism given in the main text is neutral with respect to whether 'morality' is thought of primarily as an 'individual' or as a 'social' construction. Many recent discussions of constructivism (such as the so-called 'neo-Kantian' and 'neo-Humean' versions of constructivism discussed in Lenman & Shemmer 2012) seem to have followed the former route. For various reasons I am not able to go into at greater length here, I favor the latter route.

10 I am assuming here that some cases of fear are accurate responses to a social world that is, in fact, dangerous. The same goes for other more or less 'thick' descriptions that a constructivist account of moral thought will admit as true, or otherwise correct, descriptions of social reality, but which moral skeptics (and potentially some moral realists) are likely to classify as embodying potentially dubious metaphysical claims.

11 There are interesting dangers lurking here, as manifested (for example) by the power of religious demagogues, political propagandists and persuasive advertisers. Exploring these dangers is a topic for another occasion. For the purposes of illustration, the reader might observe the affective content of virtually any minimally sophisticated online interface, or the increasing use (and apparent effectiveness) in the use of social media in mainstream politics.

12 I am very grateful to Diego Machuca, Bart Streumer and Folke Tersman for their comments on a previous draft of this chapter.

References

Albertzart, M. 2014. *Moral Principles*. London: Bloomsbury.

Blackburn, S. 1985. *Spreading the Word*. Oxford: Oxford University Press.

———. 1998. *Ruling Passions*. Oxford: Oxford University Press.

Constantinescu, C. 2014. "Moral Vagueness: A Dilemma for Non-Naturalism," *Oxford Studies in Metaethics* 9: 152–185.

Dworkin, R. 2011. *Justice for Hedgehogs*. Cambridge, MA: Harvard University Press.

Finlay, S. 2014. *Confusion of Tongues: A Theory of Normative Language*. Oxford: Oxford University Press.

Gert, J. 2012. *Normative Bedrock: Response-Dependence, Rationality and Reasons*. Oxford: Oxford University Press.

Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.

Kail, P. 2007. *Projection and Realism in Hume's Philosophy*. Oxford: Oxford University Press.

Korsgaard, C. 2009. *Self-Constitution*. Oxford: Oxford University Press.

Kramer, M. 2010. *Moral Realism as a Moral Doctrine*. Oxford: Blackwell Wiley.

Lenman, J., & Shemmer, J. (eds.). 2012. *Constructivism in Practical Philosophy*. Oxford: Oxford University Press.

Lillehammer, H. 2003. "Debunking Morality: Evolutionary Naturalism and Moral Error Theory," *Biology and Philosophy* 18: 567–581.

———. 2004. "Moral Error Theory," *Proceedings of the Aristotelian Society* 104: 94–111.

———. 2011. "Constructivism and Error Theory." In C. Miller (ed.), *The Bloomsbury Companion to Ethics*, 55–76. London: Bloomsbury.

———. 2013. "A Distinction without a Difference? Good Advice for Moral Error Theorists," *Ratio* 26: 373–390.

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

McDowell, J. 1998. *Mind, Value and Reality*. Cambridge: Harvard University Press.

Olson, J. 2014. *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press.

Price, H. 2011. *Naturalism without Mirrors*. Oxford: Oxford University Press.

Putnam, H. 2004. *Ethics without Ontology*. Cambridge, MA: Harvard University Press.

Street, S. 2006. "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127: 109–166.

Wiggins, D. 1998. *Needs, Values, Truth*, 3rd edition. Oxford: Blackwell.

Williams, B. 1981. *Moral Luck*. Cambridge: Cambridge University Press.

———. 1985. *Ethics and the Limits of Philosophy*. London: Fontana.

3 Error Theory, Relaxation and Inferentialism

Christine Tiefensee

1. Introduction

Is it possible to devise a coherent form of external skepticism about the moral? Traditionally, there has been little doubt that we should answer this question in the affirmative. After all, when declaring that all moral judgments are false because there are no moral properties, error theorists have repeatedly stressed that they are putting forward a thesis *about* the moral domain, not one *within* it: they are, to use Mackie's famous formulation, making a meta-physical claim about the "fabric of the world" (Mackie 1977: 15). However, this understanding has lately come under pressure from various directions. Ronald Dworkin (1996; 2011) was one of the first to challenge the possibility of defending external skepticism about the moral, arguing that all skeptical stances about the existence of moral properties presuppose the truth of some positive moral judgment and must thus be understood as varieties of *internal* skepticism. Minimalist conceptions of truth, fact and property, which have recently stirred much debate in the wake of Jamie Dreier's (2004) problem of creeping minimalism, appear to suggest a similar conclusion by entailing that there is no substantive difference between asserting a *moral* claim, such as "Stealing is wrong," and ascribing *truth* to this claim by asserting "It is true that stealing is wrong." At the same time, so-called 'relaxed' moral real-ists such as Tim Scanlon (2014) stress that, contrary to philosophical lore, their claims about moral ontology must not be read as robustly metaphysi-cal theses, but as *moral* commitments that are to be assessed on grounds of domain-internal, moral standards. But if relaxed moral realists' endorsement of moral truths' existence is now to be interpreted as a moral position, then surely error theorists' rejection of these very truths must be just as moral. If sound, this relaxed, non-metaphysical reading of moral ontology would spell serious trouble for error theories: were their claim that there are no moral truths itself moral, error theories would not merely lose their robustly meta-physical status, but would be outright incoherent.

   Although I will present considerations that aim to bestow some plausibil-ity on these recent developments towards relaxed, non-metaphysical read-ings of moral truth claims, my primary concern here will not be whether or

not we should indeed relax about moral ontology. Rather, I will investigate if, and if so how, a coherent, non-moral reading of error theories could be available even if it were agreed that claims about the (non-) existence of moral properties are indeed moral. After all, much has been written on how minimalism's non-metaphysical take on truth, fact and property may affect the debate between moral realists and expressivists; far less thought has been devoted to its implications for our understanding of error theories.

In this contribution, I seek to address this imbalance. To lay my cards on the table, though, I do so even though I am not an error theorist myself. Yet, whilst I believe that error theories should be rejected, I also believe that it is of general interest whether or not external skepticism about the moral is at all possible. I will answer this question in the positive: a coherent form of external error theories can indeed be devised. However, this form no longer makes any reference to the alleged falsity of moral judgments and the non-existence of moral properties. Instead, it concerns a specifically inferentialist construal which suggests that error theorists should abstain from *any* claims about moral ontology and focus exclusively on claims about the inferential role of moral vocabulary. Consequently, whereas error theories have traditionally been taken to consist of a *conceptual* component, stating that moral discourse is centrally committed to some thesis *T*, and an *ontological* component, showing this thesis *T* to be false (Joyce 2001: 5), that inferentialist construal will comprise only the conceptual element and abandon all ontological claims.

As I will show, this suggestion affords a number of important advantages. However, it also comes at a cost, in that it might change not only the letter, but also the spirit of traditional error theories. As such, this essay can be understood as offering error theorists a suggestion as to how the non-moral, external status of their position might be saved. Whether or not this proposal is one that error theorists would want to accept remains for them to decide.

The next section seeks to motivate a relaxed, non-metaphysical reading of moral ontology. As hinted above, the considerations presented will not construct a conclusive case for this relaxed stance, but are merely intended to provide sufficient grounds for making this moral reading plausible.[1] Based on an error-theoretic counterargument to this relaxed interpretation, Section 3 will lay the foundations for developing a form of skepticism which qualifies as external even if we relax about moral ontology. Section 4 will spell out this inferentialist conception of error theories in greater detail. Section 5 will conclude by considering what error theorists might gain, but also what they may lose, by adopting this modified understanding of their position.

For convenience's sake, I will help myself to some stipulated terminology. The term 'ontological' will be applied to theses about the existence and nature of truths, facts and properties. As such, 'moral ontology' will comprise claims about the existence and nature of moral truths, facts and properties. 'Metaphysical', in turn, will be used more narrowly, in that I

will reserve it for views that hold the truth of claims about moral ontology to be grounded on non-moral, metaphysical considerations. As such, I will understand metaphysical approaches as presupposing that there is a domain-neutral understanding for the concept of 'existence' as well as general existence conditions which apply equally across all domains. The relaxed position will then be taken to deny that such domain-neutral senses and criteria of existence are tenable. This essay thus presupposes a Carnapian distinction between *internal* questions that are asked within a framework or domain, and *external* questions that are posed about this domain,[2] where moral considerations count as domain-internal and metaphysical considerations as domain-external.

Despite the dislike that many of them share about this term, those advocating a moral, non-metaphysical reading of moral ontology will be called 'relaxed philosophers'. Amongst this relaxed camp, we can find a variety of different philosophers who, despite their agreement on moral interpretations of moral ontology, somewhat surprisingly differ not only on how and why to relax about moral truths, but also on most other philosophical questions.[3] Although I will follow Scanlon's (2014) account most closely, my interpretation of relaxed realism will borrow from all these different approaches without suggesting that every relaxed philosopher will agree with everything I have to say on moral readings of moral ontology. Finally, in line with the specific focus of this volume on moral skepticism, I will understand error theories locally, i.e., as encompassing the moral domain but not normative discourse more widely.[4] Arguably, the difference between moral and normative error theories impacts on how far-reaching relaxed critiques of error theories are. Whenever it does so, it will be indicated in the text.

2.   Relaxing About Moral Ontology

Assume, then, that you hear philosophers asserting claims such as "There are no moral properties," "Moral facts are *sui generis*," or "Moral properties are mind-independent." What kinds of claims are those philosophers propounding when defending these theses? According to one very plausible reading, they are putting forward metaphysical claims, which assess the ontological foundation of the moral from a standpoint *outside* of the moral domain. Given this very intuitive categorization, why believe that claims about moral ontology are not metaphysical, but moral? Put differently, why leave metaphysics behind and relax about moral ontology?

Relaxation is triggered by a diverse array of impulses—such as Hume's Law, interpretations of metaethical theses, or minimalism about truth, fact and property—and all may have their role to play when rejecting metaphysical approaches to moral ontology. Its main driver, though, concerns the way in which we are supposed to approach the very basic ontological question "What exists?" Following our philosophical training or, maybe more charitably, our intuitive inclination, we may feel disposed to tackle this question

by searching for some general criterion of existence which applies across all domains and settles whether or not a thing exists *irrespective* of the subject area in question. For instance, assume that this criterion proclaimed that existence is conditioned on causal efficacy. If so, do genes exist? Yes, they do, because genes causally impact on how we look and how likely it is for us to contract cancer, say. Do numbers exist? No, they do not, because numbers are not causally efficacious. Or imagine that existence were tied to explanatory indispensability. If so, are there electrons? Yes, there are, because electrons are explanatorily indispensable. Are there moral properties? No, there are not, because we can do without them in our best explanations (or so many philosophers argue). According to relaxed philosophers, despite such potential philosophical inclinations to the contrary, this general, domain-neutral approach to existence questions is fundamentally misguided: there are no general, substantive conditions of existence that would apply across all subject domains. Instead, existence questions can only ever be answered on grounds of the domain-*internal* standards that govern the respective subject matter. In order to justify the claim that electrons exist, say, the requirements of science will demand that electrons are causally and explanatorily efficacious. When determining whether or not there exists a Baroque age, though, we need not consult science, but aesthetic theories, whereas it is moral considerations on grounds of which we must answer questions about the existence of, say, virtue. As Scanlon (2014: 25) puts it, we "make claims expressed by the existential quantifier in many domains, but what is required to justify any existential claim, and what follows from such a claim, varies, depending on the kind of thing that is claimed to exist." Furthermore, since the notion of existence is closely related to those of truth and fact, we can add that just as existence questions do not require domain-neutral, metaphysical enquiries, nor do questions of truth and fact: no matter whether we ask if an even prime number *exists*, or if it is *true* that there is some prime number which is even, or if it is a mathematical *fact* that there is an even prime number, we are thrown back to the *mathematical* question of whether or not some prime number is even.[5] At no stage do metaphysical thoughts enter into these considerations.

Importantly, relaxing about ontology does not entail that there are no important ontological questions to be addressed. It does entail, though, that these questions are domain-specific, and as such can be answered neither on domain-neutral grounds nor on the basis of the standards of some *other* domain.[6] Just as importantly, once these domain-specific standards entail that virtue exists, or that claims about oceans and numbers are true, there is no *further* question as to whether virtue, oceans and numbers *really* exist and whether claims about them are *really* true. The very general existential question "What exists?" then, is parasitic on specific, domain-internal existential questions such as "Are there numbers?" "Are there tables?" or "Are there reasons?": we can answer the former only on grounds of answering the latter. Without the latter, we have no grip on the former.

Let us return to moral ontology, then, and consider a claim such as "It is a fact that happiness is good." As has just been mentioned, this claim could just as easily be expressed by using the formulations "It is true that happiness is good," "Happiness possesses the moral property of goodness," or simply "Happiness is good": given non-metaphysical notions of truth, fact and property, we can slide seamlessly between these four claims without adding or subtracting any metaphysical content. How do we find out whether or not we are justified in committing ourselves to the goodness of happiness? The way *not* to proceed is to start asking questions about the metaphysical nature of the property that is ascribed to happiness. That is, we should not declare that we would be justified in believing in the existence of goodness only if this seemingly 'queer' property featured in our best causal explanations or could be placed within a natural world, or if there were some metaphysical relation 'in the world', as it is often put, that would make it true or explain the goodness of happiness. Doing so would amount to imposing illegitimately the standards of the natural domain to the moral domain, and thus misunderstand what it takes for moral ontological commitment to be justified. What we need to do instead is to engage in moral inquiry and consult our moral theories to find out whether or not happiness is indeed good. Importantly, if these moral considerations show that happiness is indeed good, this *settles* that we are justified in committing ourselves to the goodness of happiness;[7] to ask for further, metaphysical support for this ontological commitment would be to misconstrue what is required to justify claims such as "Goodness exists" or "Happiness has the property of goodness."

Hence, if relaxed philosophers are right, metaphysically driven error theories are mistaken: justification of moral ontological commitment does *not* depend on domain-external, metaphysical considerations. Error theorists' skepticism is, therefore, either misguided if understood as an external position assessing moral ontological commitment on grounds of domain-external, metaphysical standards, or must be re-interpreted as a form of internal skepticism that operates from within the moral domain. Either way, relaxing about moral ontology leaves external readings of error theories in an awkward dilemma: its first horn pulls the rug from under error-theoretic attempts to reject moral truths, facts and properties on grounds of alleged metaphysical 'queerness' by showing that such metaphysical considerations do not apply to moral discourse; the second horn presupposes the truth of some moral consideration and thus renders them incoherent.[8]

Now, I appreciate that these previous paragraphs *present* relaxed views on existence, truth and fact, but do not *argue* for them. As such, they will hardly convince those philosophers who have already taken the firm stance that this relaxed approach must be rejected. However, as providing further arguments for relaxation would require us to involve ourselves in metaontological debate and thus take us too far afield, let me follow Amie Thomasson (2007: 5) in appealing to those who are still somewhat neutral and open-minded about these questions. That is, let me ask the admittedly leading

question: if you are inclined to endorse statements such as "It is wrong to read others' diaries" or "I ought to spend more time with my grannie," and if you believe that the correct moral standards support claims such as these, what *more* would you possibly require to also endorse the ontological statements that there is the property of wrongness, or that this reason exists? If, as I hope that you would do, you answer "Nothing," you take a relaxed stance on moral ontology.

3. Morality, Negations and the Opening of a Different Avenue

However, an important objection to this relaxed project will most certainly have entered error theorists' mind by now. This relies on a move which has gained so much support amongst error theorists in recent years that it cannot be ignored here.[9] Interestingly, advocates of this move agree with relaxed philosophers that statements endorsing the existence of moral truths and facts are themselves moral; they might even agree that moral realism, proclaiming the existence of such moral truths and properties, is itself a moral position. At the same time, they contradict relaxed philosophers by declaring that *denying* the existence of moral truths and facts is *not* moral, just as the statement "Moral realism is false" does not express a moral judgment (Streumer 2017: ch. 8). This arguably rather surprising move is built on two key components. The first disputes that moral discourse is closed under negation. That is, although the claim

(M)   Making others unhappy is wrong

is moral—just as the corresponding claims "(M) is true," "(M) states a moral fact," or "Generating unhappiness possesses the moral property of wrongness" are moral—its negation

(¬M) Making others unhappy is not wrong,

together with the corresponding claims that (M) is false, or that there is no moral fact that making others unhappy is wrong, or that generating unhappiness does not possess the property of wrongness, are *not* moral.

Its second component provides the explanation of why (¬M) is not moral: for a claim to qualify as such, it must conceptually entail that something satisfies a moral predicate (Streumer 2017) or, put differently, that some moral property is ascribed to some object of evaluation. Of course, (M) satisfies this requirement and is, therefore, moral; however, negations such as (¬M) do not. And since error theorists' core thesis *just is* that there are no moral truths or properties, and thus that no moral predicate is ever satisfied, we can conclude that contrary to relaxed philosophers' claim, error theories are not moral *even if* moral realism and the endorsement of moral properties' existence were indeed moral.[10]

As it stands, I do not believe that this error-theoretic move succeeds. However, it might still point error theorists in the right direction. Before coming to these positive implications, though, let us start with the negative part and ask why this retort is not strong enough to refute relaxed philosophers. One way to show as much would be to question the claim that (¬M) does not conceptually entail that some moral predicate is satisfied. After all, does (¬M) not conceptually imply that making others unhappy is permissible, and thus that there is at least one moral property which is indeed instantiated? Error theorists have recently started to block this train of thought by arguing that whilst 'not wrong' *conversationally implicates* 'permissible', no *conceptual* entailment relations hold between these predicates. If so, (¬M) would implicate, but not conceptually entail, that the predicate 'permissible' is satisfied, and would thus retain its non-moral classification.

Although I find this appeal to conversational implicatures problematic, I will not join the controversy about conceptual entailments between 'not wrong' and 'permissible' here. Instead, let me hint at a second way to attack this error-theoretic move, which questions that in order for some claim to qualify as moral, it must entail satisfaction of a moral predicate or ascription of some moral property. To motivate these qualms, take the statement

(U)   *X* is impermissible if *X* generates less utility than some alternative *Y*.

I assume that many—although certainly not all—would agree that (U) is at least very naturally read as a moral claim. Yet, (U) itself does not ascribe a moral property, nor conceptually entail that some moral predicate is satisfied. Consequently, it does not qualify as moral on the error theorist's count. Or consider the example

(¬E) Phlogiston does not exist.

Again, I presume that many would agree that (¬E) is an empirical claim. Assuming, though, that the error theorist's categorization criterion applies *tout court*, in that for some statement to count as *D*, it must conceptually entail that some *D*-predicate is satisfied, (¬E) would fail to qualify as empirical as it does not ascribe an empirical property, nor conceptually entail that some other empirical predicate is satisfied. Of course, error theorists might be willing to accept these categorizations of (U) and (¬E) as not moral and not empirical, respectively.[11] However, given their counter-intuitive flavor, they would do so at some theoretical cost.

Assume, then, that we wanted to classify (¬E) as empirical although it does not ascribe any empirical property. Since the error theorists' preferred categorization criterion cannot help us in this respect, it is clear that we must replace it with some other criterion. Here is one such alternative: (¬E) is empirical because its truth is determined by empirical methods and scientific standards. That is, (¬E) has empirical status because it is based on considerations such as the following: its truth must be examined on grounds

of observations and experiments; this includes appeal to scientific theories; for our belief in phlogiston's existence to be justified, phlogiston would have to be explanatorily indispensable and causally efficacious; (¬E) can conflict with other empirical claims, and so on. Put differently, (¬E) is very plausibly categorized as empirical because it is based on and governed by scientific norms. Could we apply the same thoughts to (U)? Arguably yes: (U) is so naturally understood as moral because its truth must be examined by moral reasoning about the link between impermissibility and utility maximization, because (U) can explain why certain actions would not be justified, because it has the ability to guide our deliberation about what to do, because it concerns our actions towards others, because it cannot be confirmed by observation and experiments, because (U) might follow from or have implications for other moral claims, because it would tell us how to act if its right-hand side were true, and so on. Arguably, then, (U) is very naturally characterized as moral because it is based on and governed by moral norms.[12]

What about (¬M)? On the one hand, the same considerations apply: contrary to error theorists' claim, (¬M) is very plausibly sorted into moral discourse because it is naturally understood as the conclusion of moral reflections, even if it does not entail satisfaction of some moral predicate. On the other hand, we should not forget that language can be used very flexibly, so that we certainly ought to beware of believing that negations such as (¬M) can *only ever* be understood as first-order, moral claims. As Plunkett and Sundell (2013) convincingly argue, claims such as (M) and (¬M) could, for instance, also be used metalinguistically, e.g., as making claims about how moral concepts ought to be used, what their presuppositions are, to which objects they can meaningfully applied, etc. That is, when declaring that a flower's turning towards the light is not virtuous, it would be rather uncharitable of me to understand you as having come to some substantive moral conclusion about this flower's moral character; instead, it would make far more sense to read your claim as communicating that we would commit a category mistake if we applied the term 'virtuous' to flowers.[13] Consequently, even though it is extremely plausible to read claims such as (¬M) as moral, it seems that they *need* not necessarily be so interpreted.

If so, we might draw two lessons from this error-theoretic counterargument to relaxed readings of moral ontology. Firstly, when determining how to categorize statements such as (¬M), it would be far too quick to conclude that (¬M) is not moral because it does not ascribe any moral property, just as it would be far too quick to read (¬M) as a moral claim because it *can* be so interpreted. Instead, in line with relaxed philosophers' general approach, we need to determine in which context—i.e., within which domain and on which grounds—(¬M) is defended and put forward.

Secondly, even if it were agreed that (¬M) can be used to convey different information and carry non-literal content, this would not necessarily disprove relaxed philosophers' take on moral ontology. Whether or not it would depends on *which* information is supposed to be communicated by

the use of (¬M). After all, if (¬M) relays information which it does not 'wear on its sleeve', it should be possible to explicate this information by use of some other claim, the nature of which is then open to debate. For instance, imagine that when you utter (M) and I utter (¬M), we should not be understood as disagreeing about the moral status that some *specific* moral theory ascribes to generating unhappiness (on this we agree), but about which moral theory correctly specifies the application conditions of 'wrong'. So let us say that by endorsing (M), you communicate something along the very rough lines of "Hedonists are right," whereas I use (¬M) to convey "Hedonists are wrong." Since these further, implicitly communicated claims are also moral, it is clear that we would *not* have left moral discourse even if (M) and (¬M) were interpreted as expressing claims about the application conditions of 'wrong'. Consequently, referring to the possibility of non-literal uses of (¬M) does not settle the debate in error theorists' favor, but simply expands it to the status of the claims that are implicitly conveyed. And if these tacitly communicated claims concerned once more error theorists' traditional theses, we would be back to square one: relaxed philosophers will object that these error-theoretic claims are either misguided or domain-internal.

Consequently, this recent error-theoretic attempt to establish non-moral categorizations of claims such as (¬M) and "There are no moral truths" is not sufficient to rebut relaxed readings of moral ontology. However, these last remarks on metalinguistic uses of (¬M) might still point error theorists in the right direction. For if error theorists could shift focus from meta*physical* concerns about the property of wrongness to meta*semantic* enquiries into moral concepts such as 'wrongness', new avenues might indeed open up, although error theorists would have to be careful about how to develop them. One way of doing so will be presented next.

4.   From Metaphysics to Metasemantics

I have briefly mentioned above that error theories generally comprise a conceptual component—for instance, proclaiming that moral discourse is committed to the existence of irreducibly normative favoring relations—and an ontological claim—in this case, that no such irreducibly normative favoring relations exist. In addition, I have not so much argued as stated that according to relaxed philosophers, error theorists' ontological thesis is either misguided if based on domain-external, metaphysical criteria which are wrongly applied to moral ontological commitment, or must be re-interpreted as a form of internal skepticism. Assume that relaxed philosophers are right: whether or not there are moral truths is a moral question. Is there some way in which error theories could still be understood as a form of *external* skepticism?

If we want to give a positive answer, the way forward seems to be clear: if error theories' ontological component is not suited to establish external status, moral ontology should be left behind and conceptual matters should

move to the fore. Indeed, this step should not feel unusual to error theorists. To start with, not all error theories are (purely) metaphysically driven: identifying some thesis that is central to a discourse, say about irreducible normativity, and then arguing that this thesis is false because irreducibly normative reasons would be metaphysically queer, is *one* way to develop an error theory, but not the *only* way. Rather, error theories can also be built on considerations of incoherence, propounding that moral thinking is flawed because it involves built-in contradictions. This is how Michael Smith (2010) interprets Mackie's claim that conceiving of moral truths as both objective and prescriptive is incoherent; arguably, it is also how Streumer's (2017: §30) argument, suggesting that moral properties would have to be both identical and non-identical to descriptive properties, could at least in part be understood.[14] Focusing on such conceptual matters is, therefore, certainly not new to error theorists, although relaxing about moral ontology will restrict how exactly this shift towards conceptual considerations can be fleshed out.

Nor is it new to modify our understanding of metaethical accounts in the wake of non-metaphysical accounts of semantic vocabulary such as truth, fact and representation. Most instructively, take the case of expressivism as an example: loosely put, pre-minimalism expressivism has traditionally been understood as a combination of the positive thesis that the meaning of moral claims is to be explained by appeal to the conative mental states expressed, and the negative thesis that there are no moral truths and facts. Given minimalist understandings of truth and fact, though, expressivists no longer want to deny that moral truths and facts exist. Post-minimalism conceptions of expressivism have thus undergone two developments. Firstly, their negative thesis has been dropped, so that expressivism is now characterized exclusively by its positive thesis about the meaning of moral vocabulary. Questions about the existence of moral truths and facts, in turn, are—fully in line with relaxed accounts of moral ontology—understood as domain-internal, moral queries. Hence, whilst expressivists such as Simon Blackburn and Allan Gibbard endorse the existence of moral truths and facts as *participants* of moral discourse, this endorsement does not form part of their distinctive *metaethical* account. Secondly, expressivism is more and more understood not as a semantic theory which specifies the meaning of moral claims, but as a *meta*semantic theory, which considers by virtue of what it is that moral claims come to possess their respective conceptual content. And this is a question which expressivists answer without using any of the referring expressions from moral discourse. Consequently, given this new, exclusive focus on their positive metasemantic thesis, neither minimalism about truth and fact nor relaxing about moral ontology undermine expressivism's status as a distinctive domain-external, metaethical theory.

Hence, if error theorists could take a page out of expressivists' book and adapt this expressivist strategy to their own cause, they should stand a good chance of securing external status for their own position, too. If so, they need to provide, firstly, a distinctive metasemantic account of moral

language which, secondly, secures the error-theoretic spirit and, thirdly, does so without employing metaphysically driven arguments about moral truths and properties.

The inferentialist account that I will present next is certainly not the only way in which these criteria could be met. However, drawing on an inferentialist theory of meaning is, I believe, particularly well-suited to do so. Accordingly, I will first give an extremely terse overview of those inferentialist elements which are relevant for my project; I will then reject an inferentialist suggestion as to how error theories could be understood; finally, I will present how I believe inferentialist error theories should be developed.

4.1.   *Inferentialism and How Not to Construe Error Theories*

According to inferentialism's central metasemantic thesis, statements possess their specific conceptual content by virtue of the *inferential role* which they assume within the practice of making statements and asking for reasons (Brandom 1994). This inferential role includes the fact that statements can license as well as be licensed by other claims; that is, they can function as premises as well as conclusions of material inferences. At the same time, these claims are not only marked by their inferential relations to other claims, but also by non-inferential relations to certain non-linguistic phenomena outside the language game. On the one hand, these phenomena concern perceptions and observations which are non-inferentially linked to observation reports such as "This cup is red" and thus provide *input* to the language game through so-called *language-entry* transitions. On the other hand, these non-linguistic phenomena pertain to actions, which are non-inferentially related to practical commitments such as "I shall spend the day with my nephews," and can thus be understood as *output* of the language game that follow from *language-exit* transitions (Brandom 1994: 234–235). Those claims which are partially characterized by language-entry transitions we can call *doxastic* or *theoretical*; those which are partially characterized by language-exit transitions we can call *practical*. Finally, different vocabularies are associated with different *functions* in that they allow us to do different things within the game of giving and asking for reasons. For instance, observation reports such as "The sun shines today" enable language-entry transitions by reporting our reliable discriminative reactions to our environment (Williams 2013). Other vocabularies have a very different, *expressive* function in that they allow us to render inferential roles explicit which would otherwise remain implicit. Take logical vocabulary as an example: although it is implicit in our inferential practices that the commitment "Humphrey is a dog" licenses and requires the commitment "Humphrey is an animal," we cannot *talk* about this inferential relation unless we are in possession of logical vocabulary that explicates this inference by stating "*If* Humphrey is a dog, *then* he is an animal." Inferentialists are, therefore, pragmatists: in

order to grasp the significance of different vocabularies, we need to understand what they allow us to *do* when using them.

Hence, if inferentialism is to help us develop a new perspective on metaethics, we need to understand metaethical accounts as carving out distinctive, competing theses about moral vocabulary within this inferentialist framework. Matthew Chrisman (2008: 353), who is arguably the most vocal supporter of inferentialist metaethics, suggests the following characterizations:[15]

> *Inferentialist Expressivism*: moral claims express practical commitments.
> *Inferentialist Moral Realism*: moral claims express theoretical commitments and some of them are true.
> *Inferentialist Error Theories*: moral claims express theoretical commitments and none of them is true.

One advantage of these inferentialist metaethical understandings is the close proximity to their traditional counterparts: inferentialist error theories differ from inferentialist expressivists in that they agree with inferentialist moral realists on how the meaning of moral vocabulary is to be explained, yet disagree with inferentialist moral realists about the existence of moral truths. As such, this proposal preserves the orthodox view that error theorists and moral realists both concur that moral claims are *purportedly representational* and that moral content is to be explained *representationally*, whilst only the latter additionally believe that these claims represent *successfully*, i.e., that some of them are true. However, as should be evident from the explanations given above, this seeming advantage turns out to be a significant disadvantage if relaxed moral ontology is presupposed as the background of our discussion. For if the existence of moral truths and successful representation continued to shoulder the burden of demarcating moral realism from error theories within this inferentialist framework, we would just find ourselves back with the relaxed philosophers' thesis that questions of successful representation are to be settled on moral grounds. If successful representation constituted the sole bone of contention between inferentialist error theorists and moral realists, therefore, they would disagree on domain-internal grounds only, but would be metaethically indistinguishable with regard to their metasemantic account of moral content. Consequently, this first inferentialist take establishes error theories neither as a distinctive metaethical position nor as a form of external skepticism, and thus fails to advance the debate.

Accordingly, if inferentialism is supposed to help secure error theories' external status, we need a more radical departure from traditional metaethical understandings than this first inferentialist suggestion can provide. This means that inferentialist error theories must differ from moral realism not with regard to moral existential theses, but in relation to their metasemantic account of moral vocabulary. And this implies that, in order to carve

out a distinctive domain-external position, error theorists need to present metasemantic theses which, firstly, differ *both* from realist and expressivist suggestions; secondly, capture the core error-theoretic thesis that something is amiss in moral discourse; and thirdly, do so without falling back on metaphysical assumptions about moral ontology.

4.2. *Inferentialist Error Theory as External Skepticism*

When tackling this task, it is helpful to start once more with error theorists' metaethical competitors. To bring out expressivists' and moral realists' metasemantic theses more clearly, let us furthermore follow recent suggestions within the creeping minimalism debate that associate expressivism with non-representationalism, and moral realism with representationalism. Transposed into an inferentialist key, this means that both moral realists and expressivists agree that meaning is explained by the inferential role, yet disagree on what the inferential role of moral vocabulary consists in:

> *Non-Representationalism*: moral vocabulary is expressive.
> *Representationalism*: moral vocabulary is non-expressive.

According to inferentialist expressivists, moral vocabulary falls into the same category as the logical terms mentioned above: it makes inferential relations *explicit* which would otherwise remain implicit (Brandom 1994). At the same time, it differs from other expressive vocabularies, such as logical terms, in that it makes *specific* inferences explicit. These are practical inferences, which have doxastic commitments as their premises and practical commitments as their conclusions.[16] To elaborate, when examining the intricate web of inferential relations together with non-inferential language exit and entry transitions, inferentialist expressivists explain that we can observe that inferences from doxastic statements, such as "Visiting one's grandmother makes her happy," to practical commitments, such as "I shall visit my grandmother," are generally endorsed as materially good inferences. Moreover, we can observe that we are clearly able to act on the basis of these practical commitments, namely by visiting our grandmothers more often. However, expressivists will stress that what we can*not* do is talk about these inferential relations, unless we possess special terms to do so. These terms, inferentialist expressivists now tell us, are moral concepts: it is they which allow us to formulate statements such as "Visiting one's grandmother is good" and thus to put into language and speak about what already exists in our practices but has so far remained implicit. Obviously, this expressivist suggestion requires further clarification and could be spelt out in different ways. However, its core idea should be clear: moral vocabulary explicates practical inferences. Consequently, if there were no such inferences to be explicated, inferentialist expressivists conclude, we would not speak in moral terms.

Adopting a representationalist stance, inferentialist moral realists disagree: moral vocabulary is not expressive. As such, they submit that inferentialist expressivists wrongly describe our practices when claiming that moral vocabulary explicates what is already an implicit component of the inferentialist web. For, when we look at the vast mesh of inferences, we can observe that inferences from statements such as "Visiting one's grandmother makes her happy" to "I shall visit my grandmother" are *not* generally regarded as sound, *unless* it is also thought that we are entitled to the claim "Visiting one's grandmother is good because it makes her happy." If we take away this premise, entitlement to the inference is generally held to collapse. Hence, a statement such as "Visiting one's grandmother is good" does *not* explicate an already existing sound inference, but provides a premise without which no sound inference would exist. Hence, moral vocabulary is not expressive, or so inferentialist moral realists conclude. Instead, it is characterized by its link to language-entry transitions: rather than making explicit what is already there, it provides new input to the language game.

Again, much more would need to be said to flesh out exactly what this inferentialist construal of representationalism involves.[17] However, what these very rough overviews have hopefully shown is that inferentialist expressivists and moral realists offer very different accounts of moral vocabulary, with the former linking it to expressive functions and the latter associating it with language-entry transitions. At the same time, I have argued that, in order to secure external status and qualify as a distinctive metaethical account, inferentialist error theorists must depart from traditional error theorists more radically by rejecting not just expressivists' non-representationalism, but also moral realists' representationalism: if they kept concurring with moral realists that moral claims are purportedly representational and disagreed only with regard to successful representation, their disagreement would be domain-internal, not external. But if inferentialist error theorists are barred from adopting representationalism and non-representationalism, what is there possibly left for them to argue? Well, they could claim that *both* inferential moral realists and expressivists are mistaken in thinking that the inferential role of moral vocabulary can be coherently specified in the first place. More precisely, they could hold that:

> *Inferentialist Error Theories**: the inferential role of moral claims is incoherent.

What could this incoherence comprise? Obviously, it will be up to inferentialist error theorists to fill in the details. Still, Michael Williams's (2013) suggestion of explaining vocabularies on grounds of so-called 'EMUs'— *e*xplanations of *m*eaning in terms of *u*se—may give us a hint as to how error theorists might proceed.

EMUs comprise three different components:

(I) A *material-inferential* (intra-linguistic) component, specifying the inferential patterns in which a concept *C* stands.
(E) An *epistemological* component, detailing epistemological demands imposed on *C*-claims.
(F) A *functional* component, determining *C*'s function.

Clauses (I) and (E) capture *how* certain terms are used, and thus arguably specify their conceptual content; (F) makes explicit what they are used *for*. Accordingly, error theorists could establish incoherence in at least two different ways. Firstly, they could argue that it is impossible to provide coherent (I)- and (E)-clauses of moral vocabulary. For instance, they could seek to establish that when engaging in moral discourse, we necessarily enter commitments which are central to moral discourse but contradictory, in that entering one of these commitments is held necessarily to preclude entitlement to another. Take moral discourse and the skepticism some error theorists harbor about categorical normativity. Using inferentialist terminology, error theorists could proffer the thesis that moral discourse commits us to the following triad of incompatible claims:

(1) Moral commitments are practical commitments.
(2) Practical commitments are treated as carrying entitlement iff they form the conclusion of a practical inference which is entitlement-preserving for a *specific* interlocutor on grounds of her particular preferences or social status.
(3) Moral commitments are treated as carrying entitlement iff they form the conclusion of a practical inference which is entitlement-preserving for *any* interlocutor, irrespective of her preferences or social status.

This suggestion should sound familiar: it emulates Joyce's (2001) error-theoretic argument. Hence, if inferentialist error theorists could convincingly argue that we are indeed centrally committed to all three claims, there would be no coherent way for us to engage in moral discourse: commitment to any two of these claims necessarily precludes entitlement to the remaining third. As such, no consistent (I)- and (E)-clauses of moral vocabulary would be possible.

The second way to spell out incoherence departs from such better-known error-theoretic approaches by not targeting (I) and (E), but (F). As such, it does not point out allegedly inconsistent positions to which we commit ourselves within the language game, but questions the possibility to provide a coherent account of moral vocabulary's very function. Above, we have encountered two competing accounts of moral vocabulary's (F)-clause: inferentialist moral realists maintain that moral terms are used for language-entry moves, whereas inferentialist expressivists hold that they are used to make implicit practical inferences explicit. Inferentialist error theorists

could now submit that both are wrong: neither realists' nor expressivists' proposed (F)-clause is tenable. Let us begin with realists' representationalism, and thus the claim that moral vocabulary facilitates language-entry transitions by allowing us to adopt positions within the language-game in response to our environment. How could error theorists attack this account? They could start by pointing out that the paradigmatic example for language-entry transitions concerns observation reports such as "This cup is red": employing the term 'red' enables us to express the reactions we have when seeing things that are red. As this shows, though, language-entry transitions are crucially to do with perception and observation, and thus *causation*.[18] However, when relaxing about moral ontology, relaxed philosophers themselves have argued that causal considerations are *inappropriate* within the moral context: to think otherwise, they have explained, is to misapply requirements that are adequate within the natural domain to the moral domain. Accordingly, taking relaxed philosophers at their word, inferentialist error theorists could smartly highlight that if relaxed philosophers are right in claiming that causal considerations do *not* apply within the moral domain, and if the language-entry transitions featuring in moral realists' suggested (F)-clause *presuppose* causal relations, then their own relaxed stance on moral ontology precludes the representationalist (F)-clause that moral realists suggest within their metasemantic account: without causal relations, moral vocabulary cannot facilitate language-entry transitions.[19] This leaves expressivists' non-representationalism, holding moral vocabulary to explicate what is already implicit in our inferential practices. Yet, with regard to this functionalist thesis, inferentialist error theorists can appropriate realists' criticism of expressivism by agreeing that it simply is not the case that we endorse certain inferences as good and then merely explicate these moves by using moral claims. On this, inferentialist realists are right: moral premises do not make inferences explicit, they make them sound. Hence, if it were the case that making others happy is good, we could indeed license our practical commitment to visit our grandmothers on grounds of the doxastic commitment that doing so would make them happy; otherwise, this doxastic commitment would confer no such entitlement on the practical commitment. Consequently, although some vocabularies may be expressive, inferentialist error theorists could conclude that moral vocabulary is not one of them. If so, moral vocabulary can neither be understood on the basis of language-entry transitions, as inferentialist moral realists would claim, nor on grounds of expressive functions, as inferentialist expressivists would have it. As such, the function of moral vocabulary remains obscure: there is no (F)-clause that could coherently specify it.

Of course, I do not mean to suggest here that either of these error-theoretic strands of thought is sufficiently spelled out. Nor do I want to claim that either argument is successful: as I said at the outset, I am not an error theorist. However, *if* they could be made to work, we would have found a coherent form of external skepticism. Firstly, this second take on inferentialist error

theories differs both from realist and expressivist accounts of moral language and assumes, therefore, a *distinctive* position on a domain-*external*, metasemantic question. Secondly, it captures the characteristic error-theoretic thrust that moral discourse is fundamentally *flawed*, in that we cannot provide a coherent account of moral vocabulary. Hence, whereas non-inferentialist error theorists may have claimed that the idea of moral properties is incoherent in that such properties would have to be both objective and prescriptive, inferentialist error theorists maintain that moral vocabulary is defective since no consistent account of its inferentialist role is available. As a by-product, this revised understanding of inferentialist error theories thus also attacks a crucial premise of the relaxed approach to moral ontology, namely that moral discourse is a well-disciplined domain that is governed by coherent standards, which provide the grounds on which the truth-values of moral existential claims are assessed. By adopting the inferentialist approach suggested here, error theorists argue that this is false: the moral domain cannot be coherently reconstructed. Thirdly, it establishes as much on the basis of inquiries into the conceptual role of moral concepts, and thus *without* drawing on metaphysical, domain-external considerations about moral ontology: linking moral vocabulary with language-entry transitions has not been ruled out on grounds of alleged metaphysical queerness, say, but by appeal to relaxed philosophers' own take on moral ontology.[20] Nor have metaphysical considerations featured when rejecting moral terms' allegedly expressive function.

Consequently, this specific inferentialist understanding of error theories neither proffers a form of representationalism nor is defined by negative theses on successful representation. As such, it emulates expressivists' strategy by characterizing the error-theoretic position in purely metasemantic, non-ontological terms. However, in contrast to expressivists, who combine their distinctive metasemantic account of moral vocabulary with the domain-internal endorsement of moral truths and facts, inferentialist error theorists eschew *all* claims on moral ontology. Relaxed moral ontology and non-metaphysical, minimalist conceptions of truth and fact can, therefore, pose no threat to the distinctive, external status of error theories. Consequently, the suggested inferentialist understanding of error theories puts error theorists back on the metaethical map without any risk of drawing them back within the moral domain.

5. Inferentialist Error Theories: Gains and Losses

How happy should error theorists be about this inferentialist turn? Admittedly, this is not quite clear. For, whilst this inferentialist proposal comes with certain advantages, it also incurs non-negligible costs.

Starting with its advantages, most importantly for our purposes, this inferentialist construal secures error theories' status as a distinctive form of external skepticism even if it were granted to relaxed philosophers that moral ontology is itself moral. Moreover, despite no longer mentioning moral truths and properties, nor successful or purported representation—indeed, despite

not even *using* moral vocabulary—this characterization aims to retain the spirit, albeit not the letter, of error theories by showing that moral discourse is fundamentally flawed. Finally, by eschewing all theses about moral truths, this re-construal of error theories also delivers the positive side-effect that error theorists no longer need to worry about the charge that their position cannot be consistently formulated. That is, whereas incoherence loomed with regard to traditional definitions in terms of the non-existence of moral truths, which lead either to error-theoretic theses' being limited to positive moral propositions only or to the thoughts on implicatures briefly discussed above, no such problems arise with regard to the suggested inferentialist understanding.

Turning to its costs, though, it is clear that although this inferentialist characterization seeks to be true to error theories' spirit, it cannot capture it entirely. After all, error theories' distinctive claim has traditionally been that although moral assertions *purport* to represent moral reality, they *fail* to do so because there is simply nothing to represent. This was, after all, what was supposed to distinguish error theories both from moral realism and expressivism. The suggested inferentialist construal leaves no space for this claim, although it does retain error theorists' thesis that moral discourse is in some way defective. Whether or not this is sufficient to preserve enough of the error-theoretic spirit is only for error theorists to decide. For those who do not want to content themselves with this clipped error-theoretic spirit, yet want to save error theories as a form of external skepticism, the task seems clear: either they must present an alternative to this inferentialist suggestion which can achieve everything that the inferentialist proposal manages to do *and* retain more of the error-theoretic spirit; or they must reject relaxed, moral readings of moral ontology.[21]

Notes

1 Given this focus on relaxed moral realism, I will not consider how error theorists could or would seek to refute non-relaxed, metaphysical versions of moral realism.
2 Yet, it does so without endorsing Carnap's (1950) own take on such an internal/external distinction.
3 These include Thomas Nagel (1986), Ronald Dworkin (1996; 2011), Simon Blackburn (1998), Matthew Kramer (2009), Derek Parfit (2011) and Thomas Scanlon (2014).
4 I do so despite generally agreeing with Streumer (2017), and thus disagreeing with Mackie (1977), Joyce (2001) and Olson (2014), that error theories cannot be understood as being confined to the moral domain only.
5 This obviously presupposes minimalist notions of truth and fact, according to which facts are no more than true propositions, where the proposition that p is true iff p. Compare also Blackburn's (1998: 78) appeal to Ramsey's ladder: "Because of . . . minimalism we can have for free what look[s] like a ladder of philosophical ascent: 'p', 'it is true that p', 'it is really and truly a fact that p'. . ., for none of these terms, in Ramsey's view, marks an addition to the original judgement." Some metaethicists have argued that although there are such

minimalist notions of truth, fact, property and existence, these are neither the only possible understandings of these concepts nor legitimate interpretations as far as ontological matters are concerned (cf. Horgan & Timmons 2015).

6 This leaves out a key qualification that Scanlon (2014: 19) proposes, namely, that some *D*-proposition *p* is true iff *D*-specific standards entail *p*'s truth *and p* does not conflict with some other domain *E*. This 'no-conflict clause' is important, as it does allow for the relevance of domain-external standards for *p*'s truth if domain *E* has jurisdiction over *p*'s presuppositions. As such, a relaxed realist such as Scanlon does not just argue that the truth of moral claims is settled by moral standards, but also that (pure) moral claims, such as "Promoting happiness is good," have no problematic presuppositions within any other domain. This is why, as I explain below, ontological and epistemological queerness arguments are counted as misguided if they mistakenly assume that moral discourse does possess such presuppositions which would make moral properties queer. Without arguing this point here, this 'no-conflict clause' helps relaxed realists to deal with objections of ontological proliferation, e.g., with regard to the existence of witches or magical elves. But it also caters for local error theories, such as those rejecting moral but not normative truths: in this case, it is the normative domain which supposedly shows that the presuppositions of the moral domain are false (cf. nn. 7 and 8). In this context, Joyce's (2001: 45–49) remarks on Carnap are particularly interesting, as they appear to accept that external skepticism about the moral domain presupposes the conceptual and normative framework of the domain of practical rationality.

7 At this point a caveat is in order: it might be argued that even the truth of a claim such as "Happiness is good" is not settled by moral considerations alone, as it also has the *normative* presupposition that there are reasons which speak for happiness in a way that is required for this claim to be true, i.e., objective, categorical or irreducibly normative reasons. I return to this thought in note 8.

8 In the previous note, I indicated that the truth of a claim such as "Happiness is good" arguably has the *normative* presupposition that there are objective reasons. This, in turn, allows local error theorists such as Mackie (1977), Joyce (2001) and Olson (2014) to submit that this moral claim is false because its normative presupposition is false: there are no such reasons. If tenable, this would indeed be a form of external, albeit normative, skepticism that even relaxed philosophers would recognize as such. However, they would *reject* this form of external skepticism as soon as it is based on the *metaphysical* objection that such objective reasons would be "metaphysically mysterious" (Olson 2014: 136). Accordingly, although error theories which encompass the moral but not the normative could qualify as a form of external skepticism about the moral, the way in which they are generally developed also mistakenly assumes that moral claims have false *metaphysical* presuppositions and thus succumb to the first horn. Still, the dilemma presented here admittedly is stronger in the case of normative error theories than in that of local, moral error theories.

9 In somewhat different versions, it can be found in Pigden (2007), Olson (2014) and Streumer (2017).

10 Note that, if successful, this move would render it unnecessary to limit error theories' thesis to positive moral claims only, as for instance suggested by Sinnott-Armstrong (2006). That is, error theorists could keep proclaiming that *all* moral claims are false without risking incoherence, as it would be fully consistent to endorse the truth of the non-moral claim (¬M) whilst proclaiming the falsity of the moral claim (M). This is a significant advantage over interpretations which understand error-theoretic theses as being limited to positive moral claims only, as there is, I believe, no principled way to distinguish between positive and negative moral statements.

11 If, in turn, they wanted to categorize (¬E) as empirical but (¬U) and (¬M) as non-moral, they would have to introduce different categorization criteria for different domains, only some of which prescribed satisfaction of some D-predicate. How plausible this move would be cannot be considered here.

12 How do we know which standards are moral and, more generally, what characterizes the moral domain? This is a very difficult question which relaxed realists cannot shirk from addressing. However, as it would take us too far afield, I will, admittedly disappointingly, continue to rely on an intuitive understanding of the moral domain here.

13 Some relaxed realists, possibly such as Kramer (2009), might point out that identification of category mistakes also relies on moral considerations, in that it is they which show why flowers are not suitable objects for moral assessment. I prefer reading category mistakes in conceptual terms, in that someone who sincerely applies terms such as 'virtuous' to flowers is not a competent user of moral language.

14 However, in contrast to the inferentialist account to be developed shortly, Streumer's account does rely on important metaphysical considerations (cf. n. 20 below).

15 This is a slight modification of Chrisman's (2008) proposal, in that Chrisman considers ethical statements and offers a slightly different account of theoretical and practical commitment. In a later paper, Chrisman (2011) proposes a suggestion based on explanatory considerations. For more details, see Tiefensee (2016).

16 For simplicity's sake, I follow Brandom (1994) here. For an alternative inferentialist account, see Chrisman (2016).

17 This is by no means an easy feat. One possible suggestion, which has recently gained some currency, suggests that this inferentialist form of representationalism may involve explaining moral vocabulary on the basis of our discriminative reactions to our moral environment, or maybe on grounds of certain explanatory theses (Chrisman 2011). Partly for the reasons mentioned below, I believe that it is extremely difficult—if not impossible!—for relaxed realists to be representationalists. Hence, it might well be the case that moral realists face a choice: either endorse representationalism and accept a robust form of realism, or relax about moral truths and defend non-representationalism.

18 Indeed, according to inferentialists, it is this that bestows *empirical* content to observation reports (Brandom 1994: 234).

19 Do relaxed moral realists have to provide a metasemantic account of moral language? As he has indicated in personal conversation, Scanlon does not appear to think so. However, I believe that this position is not tenable. Without arguing this point here, I will say only that relaxed moral realists need to draw on metasemantic theses to deal with several challenges to their position, such as those having to do with domain-individuation and ontological proliferation.

20 Arguably, in this respect it differs from Streumer's (2017: ch. 2) incoherence argument, which is partly based on metaphysical considerations about property-identity.

21 My thanks to Terry Horgan, Diego Machuca and Bart Streumer for their comments on an earlier version of this chapter.

References

Blackburn, S. 1998. *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Clarendon Press.

Brandom, R. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.

Carnap, R. 1950. "Empiricism, Semantics, and Ontology," *Revue Internationale de Philosophie* 4: 20–40.

Chrisman, M. 2008. "Expressivism, Inferentialism, and Saving the Debate," *Philosophy and Phenomenological Research* 77: 334–358.

———. 2011. "Expressivism, Inferentialism and the Theory of Meaning." In M. Brady (ed.), *New Waves in Metaethics*, 103–125. London: Palgrave Macmillan.

———. 2016. *The Meaning of 'Ought': Beyond Descriptivism and Expressivism in Metaethics*. New York: Oxford University Press.

Dreier, J. 2004. "Metaethics and the Problem of Creeping Minimalism," *Philosophical Perspectives* 18: 23–44.

Dworkin, R. 1996. "Objectivity and Truth: You'd Better Believe It," *Philosophy and Public Affairs* 25: 87–139.

———. 2011. *Justice for Hedgehogs*. Cambridge, MA: Belknap Press.

Horgan, T., & Timmons, M. 2015. "Modest Quasi-Realism and the Problem of Deep Moral Error." In R. Johnson & M. Smith (eds.), *Passions and Projections: Themes from the Philosophy of Simon Blackburn*, 190–209. Oxford: Oxford University Press.

Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.

Kramer, M. 2009. *Moral Realism as a Moral Doctrine*. Chichester: Wiley-Blackwell.

Mackie, J. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin.

Nagel, T. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

Olson, J. 2014. *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press.

Parfit, D. 2011. *On What Matters*, Volume 2. Oxford: Oxford University Press.

Pigden, C. 2007. "Nihilism, Nietzsche and the Doppelganger Problem," *Ethical Theory and Moral Practice* 10: 441–456.

Plunkett, D., & Sundell, T. 2013. "Disagreement and the Semantics of Normative and Evaluative Terms," *Philosophers' Imprint* 13/23: 1–37.

Scanlon, T. 2014. *Being Realistic about Reasons*. Oxford: Oxford University Press.

Sinnott-Armstrong, W. 2006. *Moral Skepticisms*. Oxford: Oxford University Press.

Smith, M. 2010. "Dworkin on External Skepticism," *Boston Law Review* 90: 509–520.

Streumer, B. 2017. *Unbelievable Errors*. Oxford: Oxford University Press.

Thomasson, A. 2007. *Ordinary Objects*. Oxford: Oxford University Press.

Tiefensee, C. 2016. "Inferentialist Metaethics, Bifurcations and Ontological Commitment," *Philosophical Studies* 173: 2437–2459.

Williams, M. 2013. "How Pragmatists Can Be Local Expressivists." In H. Price, S. Blackburn, R. Brandom, P. Horwich, & M. Williams (eds.), *Expressivism, Pragmatism and Representationalism*, 128–144. Cambridge: Cambridge University Press.

4 Why We Really Cannot Believe the Error Theory

Bart Streumer

1. Introduction

According to the error theory, normative judgments are beliefs that ascribe normative properties, but these properties do not exist. I have argued elsewhere that we cannot believe this theory. Several philosophers have replied to this argument. In this essay, I will reply to their replies.[1]

This may seem tedious, but I think it matters. For I have also argued that our inability to believe the error theory makes this theory more likely to be true, by undermining objections to the theory, by making it harder to reject my arguments for the theory, and by undermining revisionary alternatives to the theory (Streumer 2013a; 2016a; 2017: sects. 72–74). And it matters whether the error theory is true.

The error theory I defend applies to all normative judgments. But my replies are also relevant to moral error theorists. For it is often suggested that the arguments for a moral error theory actually support an error theory about all normative judgments, and this is often taken to be an objection to moral error theories (see, e.g., Cuneo 2007). Moral error theorists can answer this objection by endorsing what I say in this essay.

2. Why We Cannot Believe the Error Theory

I will first repeat my argument.[2] If you are already familiar with it, you can skip this section and continue with Section 3.

I use the term 'belief' in such a way that at least two conditions have to be met for a person to believe that p. The first is that

(B1) A person believes that p only if this person is very confident that p.

We can distinguish *full* from *partial* belief: we fully believe that p if we are very confident that p, and we partly believe that p if we are only somewhat confident that p. I will use the term 'belief' to mean full belief.[3]

The second condition is that

(B2) A person believes that p only if this person adequately understands p.

Suppose I do not understand the general theory of relativity. If (B2) is true, I then do not believe this theory. Suppose next that a physicist tells me that the general theory of relativity is true. I may then come to believe that this theory is true. But if (B2) is true, I do not thereby come to believe the theory itself. I only come to believe the theory itself if I adequately understand it. On the other hand, if I *do* adequately understand the general theory of relativity, I cannot come to believe that this theory is true without thereby coming to believe the theory itself.[4]

We can also distinguish *explicit* from *implicit* belief: we explicitly believe that p if we currently think that p, and we implicitly believe that p if our current thoughts commit us to p, for example, by presupposing or entailing p. And we can distinguish *occurrent* from *dispositional* belief: we have an occurrent belief that p if we currently think that p, and we have a dispositional belief that p if we are disposed to think that p in certain circumstances, for example, when someone asks us whether it is the case that p.[5] I use the term 'belief' to mean explicit and occurrent full belief.

Since our concept of a belief is not entirely precise, there are different correct ways to use the term 'belief'.[6] If we want to set the bar low, we can say that a person believes that p even if this person is only somewhat confident that p or does not adequately understand p. If we want to set the bar higher, we can say that a person believes that p only if conditions (B1) and (B2) are met. And if we want to set the bar even higher, we can add further conditions: for example, we can say that a person believes that p only if this belief is stable and influences his or her actions. As I have said, I will set the bar at meeting conditions (B1) and (B2). But I take (B1) and (B2) to be partly stipulative: I take these conditions to pick out *a* correct way to use the term 'belief', not *the* correct way.

I think that if we use the term 'belief' in accordance with (B1) and (B2), two further conditions also have to be met for a person to believe that p.[7] The first of these is that

(B3) A person believes that p only if this person believes what he or she believes to be entailed by p.

To see this, suppose that Bob says:

I believe that Socrates was a man, and I believe that this entails that Socrates was a human being, but I do not believe that Socrates was a human being.

Bob may then be insincere, or may be considering whether to give up one of these beliefs, or may fail to adequately understand what he is saying. If he is insincere, he does not believe what he says he believes. If he is considering whether to give up one of these beliefs, he is no longer very confident about at least one of the things he says he believes, which means that he fails to meet condition (B1). But he may also be neither insincere nor considering whether to give up one of these beliefs. In that case, however, he is too

confused to adequately understand what he is saying, which means that he fails to meet condition (B2). In none of these cases does Bob believe what he says he believes. This suggests that if we use the term 'belief' in accordance with conditions (B1) and (B2), condition (B3) has to be met as well.

If you doubt this, this may be because you conflate (B3) with a different claim, such as:

(1) A person *partly* or *implicitly* or *dispositionally* believes that p only if this person believes what he or she believes to be entailed by p.
(2) A person believes that p only if this person believes what he or she *partly* or *implicitly* or *dispositionally* believes to be entailed by p.
(3) A person believes that p only if this person believes what is *actually* entailed by p.
(4) A person believes that p_1, that p_2, . . . and that p_n only if this person believes what he or she believes to be entailed by the conjunction of p_1, p_2, . . . and p_n.

But (B3) is a weaker claim than (1) to (4). Even if (1) to (4) are false, therefore, (B3) can be true. That is what I think is the case.[8]

The second further condition that I think has to be met if we use the term 'belief' in accordance with (B1) and (B2) is that

(B4) A person believes that p only if this person does not believe that there is no reason to believe that p.

To see this, suppose that Bob says:

> I believe that Socrates was a man, but I believe that there is no reason to believe this.

As before, Bob may then be insincere, or may be considering whether to give up one of these beliefs, or may fail to adequately understand what he is saying. If he is considering whether to give up one of these beliefs, he is no longer very confident about at least one of the things he says he believes, which means that he fails to meet condition (B1). But he may also be neither insincere nor considering whether to give up one of these beliefs. In that case, however, he is too confused to adequately understand what he is saying, which means that he fails to meet condition (B2). In none of these cases does Bob believe what he says he believes. This suggests that if we use the term 'belief' in accordance with conditions (B1) and (B2), condition (B4) has to be met as well.

As before, if you doubt this, this may be because you conflate (B4) with a different claim, such as:

(5) A person *partly* or *implicitly* or *dispositionally* believes that p only if this person does not believe that there is no reason to believe that p.

(6) A person believes that *p* only if this person does not even *partly* or *implicitly* or *dispositionally* believe that there is no reason to believe that *p*.

(7) A person believes that *p* only if this person knows *what* reason there is to believe that *p*.

(8) A person believes that *p* only if this person believes that *there is* a reason to believe that *p*.

(9) A person believes that *p* only if this person does not believe that there is no consideration that stands in an *irreducibly normative favoring relation* to this belief.

(10) A person *accepts* that *p* only if this person does not believe that there is no reason to believe that *p*.

(11) A person believes that *p* only if this person does not believe that there is no *evidence* that *p*.[9]

But as before, (B4) is a weaker claim than (5) to (11). Even if (5) to (11) are false, therefore, (B4) can be true. As before, that is what I think is the case.[10]

Suppose I am right that if we use the term 'belief' in accordance with (B1) and (B2), conditions (B3) and (B4) also have to be met for a person to believe that *p*. I think it follows from this that we cannot believe the error theory. As I have said, according to the error theory, normative judgments are beliefs that ascribe normative properties, but normative properties do not exist. And I have argued elsewhere that the property of being a reason for belief is a normative property (see Streumer 2016a; 2017: sect. 51). The error theory therefore entails that there is no reason to believe this theory. And anyone who understands the theory well enough to believe it knows that it entails this. Therefore, given that

(B3) A person believes that *p* only if this person believes what he or she believes to be entailed by *p*,

anyone who believes the error theory believes that there is no reason to believe this theory. But given that

(B4) A person believes that *p* only if this person does not believe that there is no reason to believe that *p*,

that is impossible. This means that we cannot believe the error theory.[11]

3. Are (B3) and (B4) Only True of Rational Beliefs?

As I have said, several philosophers have replied to this argument. A first reply, which is given by Hallvard Lillehammer and Niklas Möller, is that (B4) is only true of rational beliefs (Lillehammer & Möller 2015: 456). A similar claim could be made about (B3).

I agree that if a belief meets conditions (B3) and (B4), this belief is rational in a certain sense: it is closed under believed entailment, since the person

who has this belief believes what he or she believes to be entailed by this belief, and it is not believed to be unsupported, since the person who has this belief does not believe that there is no reason for this belief. But that is no objection to my argument. For as Lillehammer and Möller say, being closed under believed entailment and not being believed to be unsupported are descriptive properties, which means that being rational in this sense is also a descriptive property.[12] Since I think that if we use the term 'belief' in accordance with (B1) and (B2), conditions (B3) and (B4) also have to be met, I take this descriptive property to be partly constitutive of the mental state that this use of the term 'belief' picks out. If you deny that this descriptive property is partly constitutive of belief, I think you are setting the bar for believing that p lower than I have done: you may think, for example, that a person can believe that p even if he or she is only somewhat confident that p, or even if he or she does not understand p. I agree that if we set this bar lower, we can believe the error theory (see Streumer 2017: sect. 63).

4. Are There Counterexamples to (B4)?

Another reply to my argument is that there are counterexamples to (B4). One purported counterexample is constituted by religious beliefs. Suppose that Susan says:

> I believe that God exists, but I believe that there is no evidence for this belief.

She can then perhaps believe what she says she believes. But she may take there to be other reasons to believe that God exists: she may think that she will be denied entry to heaven if God exists but she fails to believe this, as Pascal thought, or that believing that God exists enables her to become her true self, as Kierkegaard seems to have thought.[13] For this example to be a counterexample to (B4), Susan should instead say:

> I believe that God exists, but I believe that there is no reason to believe this.

She may then seem to believe what she says he believes. But I think that if we use the term 'belief' in accordance with (B1) and (B2), Susan does not really believe this. She may instead use the term 'reason for belief' to mean evidence, or she may use the term 'belief' to mean acceptance, or she may merely mean to say that she does not know what reason there is to believe that God exists.

Lillehammer and Möller disagree. They write:

> Given some of the things that are at stake in matters of faith it is hardly surprising that someone's level of confidence in a religious proposition can vary independently of reasons they take to exist in favour of its truth.

(Lillehammer & Möller 2015: 455)

But I do not deny this. I only claim that we cannot be *very confident* in a religious proposition while believing that there is *no reason whatsoever* to believe it. They continue:

> To think otherwise is to confuse the (hopeful) belief that God exists with the belief that He has provided us with reasons to believe in His existence. Whether He either could or should provide us with such reasons is a matter of theological dispute.
>
> (Lillehammer & Möller 2015: 455)

But I do not deny this either. I only claim that we cannot believe that God exists while at the same time believing that *there is* no reason to believe that He exists, not even a reason that He has not revealed to us. If Susan thinks that God has not revealed such a reason to us, she does not believe that there is no reason to believe that God exists, but only that we do not know what this reason is.

A second purported counterexample to (B4) is constituted by self-evident beliefs. Suppose that Susan takes it to be self-evident that $1 + 1 = 2$, and suppose that she says:

> I believe that $1 + 1 = 2$, but I believe that there is no reason to believe this.

She may then also seem to believe what she says she believes. But p is self-evident if and only if adequately understanding p gives us sufficient reason to believe that p. This means that if Susan takes it to be self-evident that $1 + 1 = 2$, she does not really believe that there is no reason to believe that $1 + 1 = 2$.[14]

Marianna Bergamaschi Ganapini (2016: 528) similarly suggests that basic beliefs and beliefs in what Wittgenstein calls "hinge propositions" are counterexamples to (B4). I agree that we can have such beliefs without believing that there are reasons for them. But (B4) only says that we cannot have a belief *while* believing that there is *no* reason for it. And when foundationalists call certain beliefs "basic" and Wittgenstein calls certain propositions "hinge propositions," they do not mean that we can have such beliefs while believing that there are no reasons for them. They only mean that we can be justified in having such beliefs without basing them on other beliefs.

A third purported counterexample to (B4) is constituted by compulsive or deluded beliefs (Streumer 2013a: 197; Bergamaschi Ganapini 2016: 528). Suppose that Susan suffers from the Capgras delusion, which makes her think that her family members have been replaced by robots, and suppose that she says:

> I believe that my husband has been replaced by a robot, but I believe that there is no reason to believe this.

She may then also seem to believe what she says she believes. But I think that if we use the term 'belief' in accordance with (B1) and (B2), even people with a compulsive or deluded belief probably do not believe that there is no reason whatsoever for this belief. For example, patients who suffer from the Capgras delusion have brain damage that gives them certain abnormal experiences, and they mistakenly assume that the best explanation for these experiences is that their family members have been replaced by robots (Bortolotti 2010: 120). They therefore do seem to take there to be reasons for their deluded beliefs. Moreover, if some people with compulsive or deluded beliefs believe that there are no reasons for these beliefs, I could revise (B4) to:

> (B4*) A person believes that *p* only if this person does not believe that there is no reason to believe that *p*, unless the belief that *p* is compulsive or deluded.

Since a belief in the error theory is not compulsive or deluded in the relevant sense, this revision would not undermine my argument.

A fourth purported counterexample to (B4) is constituted by certain philosophical beliefs. Suppose that a lecture on skepticism convinces Susan that there is no reason to believe that her senses are reliable, and suppose that she says:

> I believe that my senses are reliable, but I believe that there is no reason to believe this.[15]

She may then also seem to believe what she says she believes. But I think that if we use the term 'belief' in accordance with (B1) and (B2), Susan does not really believe this. As before, she may instead use the term 'reason for belief' to mean evidence, or she may use the term 'belief' to mean acceptance, or she may merely mean to say that she does not know what reason there is to believe that her senses are reliable. Alternatively, she may have different beliefs at different times: during the lecture she may be very confident that there is no reason to believe that her senses are reliable while being only somewhat confident that her senses are reliable, and after the lecture she may again be very confident that her senses are reliable while being only somewhat confident that there is no reason to believe this (see also Lillehammer & Möller 2015: 456; Forcehimes & Talisse 2016: 851).

Alexander Hyun and Eric Sampson disagree. They think that Susan

> may be convinced by skeptical arguments that she has no reason to believe that her senses are reliable and, at the same time, believe that they are, either because (i) nature has constituted her so that she cannot help but believe that they are reliable; (ii) she thinks that she could not possibly give reasons in their defense; (iii) she is convinced that her life

will go much better if she believes that they are reliable; or (iv) all of the above.

<div align="right">(Hyun & Sampson 2014: 634)</div>

But suppose first that (i) is true: suppose that Susan cannot help but believe that her senses are reliable. In that case, her belief that her senses are reliable is, in effect, compulsive. As I have said, if compulsive beliefs are a counter-example to (B4), I can revise (B4) to (B4*) without thereby undermining my argument (see also Forcehimes & Talisse 2016: 851–852). Suppose next that (ii) is true: suppose that Susan thinks she could not possibly give rea-sons to believe that her senses are reliable. In that case, Susan does not really believe that *there is* no reason to believe that her senses are reliable, but only that she cannot say what reason there is to believe this. And (B4) only says that we cannot have a belief while believing that there is no reason for this belief, not that we cannot have a belief while being unable to say what reason there is for it. Finally, suppose that (iii) is true: suppose that Susan is convinced that her life will go much better if she believes that her senses are reliable. She then does seem to take there to be a reason to believe that her senses are reliable: namely, that her life will go much better if she believes this.[16] I therefore think that these examples fail to undermine (B4).

5. Is a Belief in the Error Theory Itself a Counterexample to (B4)?

Some philosophers think, however, that a belief in the error theory itself constitutes a counterexample to (B4). They think that since someone who believes the error theory believes that there are no reasons at all, such a per-son can have a belief while believing that there is no reason for this belief. This sounds suggestive, but I am not sure what to make of it. Of course, it is true that

(1) If we believed the error theory, (B4) would be false of us.

But the mere fact that (B4) *would* be false of us *if* we believed the error theory does not show that (B4) is *actually* false of us. Merely pointing out that (1) is true is therefore not enough to show that we can believe the error theory.[17]

It is also true that

(2) We can believe that there are no considerations that stand in an irreduc-ibly normative relation to our beliefs.

But that does not show that (B4) is false, since (B4) does not say that a person believes that *p* only if this person accepts non-reductive realism about reasons for belief. Of course, to believe an error theory about all normative judgments,

we must believe that judgments about reasons for belief ascribe a normative relation, since otherwise there would be normative judgments about which we would not believe the error theory. But we do not need to believe that this relation is *irreducibly* normative. Moreover, we do not need to believe that judgments about reasons for belief ascribe a normative relation for (B4) to be true of us, but only for it to be true that we believe the error theory.

Jonas Olson gives a more specific version of this example. He writes:

> I can . . . base my belief that the error theory is true on the argument from queerness, without judging that this argument favours my attitude of believing that the error theory is true. I can thus maintain that while there are arguments on which I base my belief that the error theory is true, there are no irreducibly normative reasons for the attitude of believing that the error theory is true. Hence we can indeed believe the error theory.
>
> (Olson 2014: 171–172)[18]

Olson is here replying to my earlier defense of (B4), in which I wrote that "reasons for belief are considerations that we *base* our beliefs on, and we cannot base a belief on a consideration without making at least an implicit normative judgment" (Streumer 2013a: 198). In response, he points out that children and non-human animals form beliefs on the basis of perception "although they presumably lack the relevant normative thoughts" (Olson 2014: 171). I agree. I therefore withdraw my earlier claim that basing a belief on a consideration involves making an implicit normative judgment.[19]

But that does not undermine my argument. For though Olson is right that

(3) A person can base a belief on a consideration *without* believing that this consideration is *a* reason for this belief,

this does not mean that

(4) A person can base a belief on a consideration *while* believing that this consideration is *no* reason for this belief.

If I perceive that the desk at which I wrote this essay is white, I will normally form the belief that this desk is white on the basis of this perceptual input without making a normative judgment. But suppose I believe that I have taken a powerful drug that makes red objects look white to me. In that case, my perceptual input will not change: the desk at which I wrote this essay will still look white to me. But I will now believe that this input is no reason to believe that this desk is white. And if I believe that this perceptual input is no reason for this belief, I will be unable to form the belief that this desk is white on the basis of this input.[20]

Similar claims apply to any other consideration that we can base our beliefs on. Suppose I believe that men are not human beings. I will then be unable to form the belief that Socrates is a human being on the basis of the consideration that Socrates is a man, since I will then believe that this consideration is no reason for this belief. Or suppose I believe that only evidence can be a reason for a belief. I will then be unable to form the belief that God exists on the basis of the consideration that I will be denied entry to heaven if God exists, since I will then believe that this consideration is no reason for this belief. Even if Olson is right that (3) is true, therefore, this does not mean that (4) is true.[21]

Bergamaschi Ganapini disagrees. To support (4), she gives the following example:

> Imagine a professor on a job committee examining the dossiers of two job applicants: Paul and Gina. They are both good candidates, but Gina is more qualified than Paul based on some 'objective standards' (e.g. number of publications). Contrary to that, however, the professor believes that the right thing to do is to hire Paul. Suppose for now that his belief is—unbeknownst to him—in part based on his belief that Gina is a woman. Suppose further that he explicitly denies that being a woman is a reason for not hiring someone who is well qualified for the job. Now if this is a possible scenario—as I believe it is—we have a situation in which someone's belief is based on a consideration that he or she does not see as a reason for that belief.
>
> (Bergamaschi Ganapini 2016: 530)[22]

There are different views about what it is for a person's belief to be based on a consideration C: this may be, for example, that C caused this belief, or that this person is disposed to revise this belief if C does not obtain, or that this person at least implicitly takes C to be a reason for this belief.[23] If the first or second view is true, this example shows that

(5) A person's belief can *be based* on C even though this person believes that C is no reason for this belief.

But there are two things the example does not show. First, it does not show that

(6) A person can *base* a belief on C while believing that C is no reason for this belief.

If a person bases a belief on C, this is something that this person actively does rather than something that merely happens to him or her. A person who bases a belief on C must therefore realize that this belief is based on C. And if this professor realizes that his belief that Paul should be hired is based

on the consideration that Gina is a woman, I think he cannot continue to have this belief while believing that this consideration is no reason for this belief. I therefore think that this example does not support (4).

Moreover, this professor surely does not believe that there is *no reason whatsoever* to believe that Paul should be hired. Instead, he presumably has certain false beliefs that he takes to be reasons for this belief: he may falsely believe, for example, that Paul's publications are better than Gina's, or that Paul is more impressive in discussion that Gina, or he may believe some other rationalization along these lines. Of course, this does not mean that Paul should actually be hired. But it does mean that this example does not undermine (B4).

6. Is There a Way in Which We Can Come to Believe the Error Theory?

Another reply to my argument is that there are specific ways in which we can come to believe the error theory. One way in which we may seem able to do this is that

(1) We can believe that judgments about reasons for belief are not norma-tive judgments, and we can then come to believe an error theory about all judgments that we take to be normative.

I agree that we can do what (1) describes. But if I am right that judgments about reasons for belief are normative judgments, we would then not believe an error theory about all judgments that are *actually* normative. Instead, we would merely believe an error theory about all judgments that *we take to be* normative. And when I say that we cannot believe the error theory, I mean that we cannot believe an error theory about all judgments that are actually normative.

A second way in which we may seem able to come to believe the error theory is that

(2) We can endorse a revisionary view about reasons for belief, and we can then replace our judgments about reasons for belief with certain descriptive beliefs, such as descriptive beliefs about evidence.

But if I am right that judgments about reasons for belief are normative judg-ments, we then would not believe an error theory about all judgments that are *currently* normative. Instead, we would merely believe an error theory about all judgments that would *then* be normative: we would believe an error theory about all judgments that are normative after we have replaced some of our normative judgments with descriptive beliefs. And when I say that we cannot believe the error theory, I mean that we cannot believe an error theory about all judgments that are currently normative.

A third way in which it we may seem able to come to believe the error theory is that

(3) We can come to believe the error theory without thereby coming to believe that there is no reason to believe this theory.

Hyun and Sampson think that we can do this, and so do Andrew Forcehimes and Robert Talisse. Both illustrate (3) with an example. Hyun and Sampson's example is as follows:

> A person might fully believe that there are no animals in the room, and hence understand this claim well enough to be in a position to believe it, but fail to believe (and hence to know) that this claim entails that there are no falcons in the room. Perhaps her thoughts simply do not turn to falcons in a way that would give rise to beliefs about them.
>
> (Hyun & Sampson 2014: 635)

And Forcehimes and Talisse's example is as follows:

> I am an arthritis-denier. I know what is typically claimed about arthritis—how it stiffens the joints, commonly occurs in wrists, fingers, and ankles, and so forth. But, because I think the elderly made up arthritis to trick the young into doing work for them, I hold that (i) when people make arthritis diagnoses they ascribe arthritis properties, and (ii) arthritis properties do not exist. Next suppose, on account of my red, swollen toe, I go to the doctor. She tells me that I have a bad case of gout. I believe her. When I arrive home, I give my partner the bad news. To my surprise, my partner explains that gout is a form of arthritis.
>
> (Forcehimes & Talisse 2016: 852)[24]

I agree that we can believe that there are no animals in the room without believing that there are no falcons in the room, and I agree that we can believe that arthritis does not exist without believing that gout does not exist. But does this show that we can do what (3) describes? Suppose that Fred is trying to believe the error theory. In other words, suppose he is trying to believe that

(4) Normative judgments are beliefs that ascribe normative properties, but these properties do not exist.

As I have said, when I say that we cannot believe the error theory, I mean that we cannot believe an error theory about all judgments that are *actually* normative. If I am right that these judgments include judgments about reasons for belief, to come to believe the error theory, Fred must come to believe that

(5) Judgments about reasons for belief are beliefs that ascribe the property of being a reason for belief, but this property does not exist.

This means that he cannot come to believe the error theory the way Forcehimes and Talisse's example suggests: he cannot come to believe the error theory by failing to realize that this theory applies to judgments about reasons for belief.

   To believe (5), Fred must believe both of its conjuncts. He must therefore believe both that

(6) Judgments about reasons for belief are beliefs that ascribe the property of being a reason for belief

and that

(7) The property of being a reason for belief does not exist.

If (5) is true, believing (7) is equivalent to believing that

(8) There are no reasons for belief.

This means that in order to believe the error theory, Fred must believe (8). And, of course, (8) entails that

(9) There is no reason to believe the error theory.

   Now suppose that Fred is trying to believe the error theory. Can he do this while temporarily ignoring the entailment from (8) to (9)? Perhaps he could if he was not explicitly thinking about the error theory. But since Fred is trying to believe the error theory, he *is* explicitly thinking about this theory. I therefore think that he also cannot come to believe the error theory the way Hyun and Sampson's example suggests: he also cannot come to believe the error theory by failing to connect (8) to the error theory. A fourth way in which we may seem able to come to believe the error theory, which is suggested by Bergamaschi Ganapini, is that

(10) We can come to believe the error theory while taking the arguments for this theory to be reasons for this belief, and once we believe the error theory we can stop taking these arguments to be reasons for this belief.
(Bergamaschi Ganapini 2016: 530)

But I think we cannot do this either. For as I have just explained, to believe the error theory, we must believe that

(8) There are no reasons for belief.

And if while trying to believe the error theory we take the arguments for this theory to be reasons for this belief, this will prevent us from coming to believe (8). It will therefore prevent us from coming to believe the error theory.

Finally, Hyun and Sampson suggest that there must be *some* way in which we can come to believe the error theory, since there seems to be a possible world in which people believe this theory. As they describe it, in this world,

> the error theory is taught to school children from an early age. On Sunday mornings, everyone gathers in large buildings in their communities where they hear readings from *Ethics: Inventing Right and Wrong*, recite the error-theory creed, and sing hymns about J. L. Mackie. In this way, children and young adults are instructed with the teachings of the error theory. There is no opposition. No one has ever heard of a different meta-ethical theory.
>
> (Hyun & Sampson 2014: 633)[25]

But reciting the "error-theory creed" and singing hymns about Mackie is not enough to believe the error theory. Do these people make normative judgments the way we do? Are their judgments about reasons for belief normative the way ours are? And if we use the term 'belief' in accordance with (B1) and (B2), do these people then nevertheless believe the error theory? It is hard to say, but I think they do not. What Hyun and Sampson are describing, I think, is merely a world in which people *seem* to believe the error theory.

7.   Can We Come Close to Believing the Error Theory?

Though I think that we cannot believe the error theory, I have also argued elsewhere that we can come close to believing this theory. I think there are at least two ways in which we can do this (Streumer 2013a).[26] First, we can believe different parts of the error theory at different times, while implicitly changing some of our other beliefs. When we consider arguments for the claim that

(1) Normative judgments are beliefs that ascribe normative properties,

we can believe (1), while at the same time failing to believe that

(2) Normative properties do not exist

and instead implicitly believing that normative properties do exist. And when we consider arguments for (2), we can believe (2), while at this time failing to believe (1) and instead implicitly believing that normative judgments are non-cognitive attitudes rather than beliefs that ascribe normative properties. Something like this happens to me when I consider the arguments I have elsewhere given for (1) and (2).[27]

Bergamaschi Ganapini asks why this does not count as coming close to "believing that the error theory is false," since "that there are normative properties and that normative judgements are non-cognitive attitudes are supposed to be claims incompatible with the error theory" (Bergamaschi Ganapini 2016: 531). I have three answers to this question. First, whereas my belief in (1) and my belief in (2) are explicit, the other beliefs I form while coming to believe (1) or (2) are merely implicit. Second, whereas my belief in (1) and my belief in (2) are based on what I take to be sound arguments, the other beliefs I form while coming to believe (1) or (2) are not based on arguments. Instead, I form these other beliefs merely to enable myself to believe (1) or (2). Third, I know that the reason why I am temporarily giving up my belief in either (1) or (2) is not that I am convinced that this part of the error theory is false, but is instead that this is the only way in which I can come to believe the other part of the theory.

Hyun and Sampson suggest that coming close to believing the error theory in this way may make us believe a moral error theory. They write that

> a way to come close to believing the error theory is to believe those theses that are parts of the error theory, and surely *moral error theory* is a part of the error theory. So if there are reasons to come close to believing the error theory, then there are reasons to believe moral error theory, and as a result our deepest and most important moral convictions are . . . threatened.
>
> (Hyun & Sampson 2014: 640)

But my arguments for (1) and (2) will only make us believe a moral error theory if we mistakenly think that these arguments fail to apply to judgments about reasons for belief. To make sure that coming close to believing the error theory in this way does not make us believe a moral error theory, we should therefore keep in mind that these arguments do in fact apply to judgments about reasons for belief.[28]

A second way in which I think we can come close to believing the error theory is by believing that there are sound arguments that together seem to show that the error theory is true. What does this mean? It does not mean believing that

(3) There are *seemingly* sound arguments for the error theory.

In other words, it does not mean believing that there are arguments for the error theory that are *prima facie* sound, but that may on closer inspection turn out to be unsound. Instead, it means believing that

(4) There are *sound* arguments that together *seem* to show that the error theory is true.

My belief in (4) is in one way similar to the beliefs we have when appearances are deceptive. When we put a stick in the water, this stick seems bent,

but we do not believe that it is actually bent. When we are traveling in the desert, there may seem to be an oasis in the distance, but we may not believe that there is actually an oasis in the distance. In a similar way, when I consider my arguments for (1) and (2), these arguments together seem to show that the error theory is true, but I do not believe that they actually show this. Instead of failing to believe this because I take that these arguments to be unsound, however, I fail to believe this because I cannot believe what these arguments seem to show.[29]

Bergamaschi Ganapini replies that if (B3) and (B4) are true, the following claims also seem true:

(B5) A person cannot believe that p entails q and that there are sound arguments that together seem to show that p without believing that there are sound arguments that together seem to show that q.

(B6) A person cannot believe that there are sound arguments that together seem to show that p while believing that there are sound arguments that together seem to show that there is no reason to believe that p.
(Bergamaschi Ganapini 2016: 532)[30]

If (B5) is true, and if anyone who understands the error theory well enough to be in a position to believe it knows that this theory entails that there are no reasons to believe it, then anyone who believes (4) also believes that

(5) There are sound arguments that together seem to show that there are no reasons for belief.

Moreover, it seems that we cannot come to believe (5) on the basis of (4) without thereby coming to believe that

(6) There are sound arguments that together seem to show that there is no reason to believe the error theory.

If so, and if (B6) is true, it follows that we cannot believe both (4) and (5) at the same time.[31] Bergamaschi Ganapini therefore denies that we can come close to believing the error theory in this second way.

I think, however, that (B6) is false. Just as believing (4) cannot give rise to a belief in the error theory because we cannot believe the error theory, believing (5) also cannot give rise to the belief that there are no reasons for belief because we cannot believe that there are no reasons for belief. For the fundamental reason why we cannot believe the error theory is that we cannot believe that there are no reasons for belief. Since believing (4) and believing (5) cannot give rise to these beliefs, I think we can believe both (4) and (5) at the same time. This would be like believing that our sense perception seems to show that there is an oasis in the distance, while at the same time believing that our map seems to show that there is no reason to believe this.

8.   Conclusion

I conclude that the replies that I have discussed fail to undermine my argument.[32] The argument may still seem to have what Forcehimes and Talisse (2016: 851) call "the scent of the illicit." But I think that what gives it this scent is not that it is unsound. Instead, I think it has this scent because it is hard to believe that there are philosophical theories that we cannot believe. Fortunately, we can believe that we cannot believe the error theory. I believe this. And if I can believe it, so can you.[33]

Notes

1   This chapter consolidates my replies to these replies in one place. It will therefore repeat some claims I made in Streumer (2016a; 2016b). It also draws on parts of chapter IX of Streumer (2017), by kind permission of Oxford University Press.

2   I first gave this argument in Streumer (2013a). For a more elaborate version of the argument, see Streumer (2017: sects. 56–57).

3   I am neutral about whether beliefs are binary or come in degrees: if they are binary, I use the term 'belief' simply to mean belief, and if they come in degrees, I use the term 'belief' to mean a high degree of belief.

4   You may think that (B2) sets the bar for believing that p too high. But note that (B2) is particularly plausible if p is the name of a theory or thesis, such as 'the general theory of relativity', 'Goldbach's conjecture', or 'the error theory'. A version of (B2) that is restricted to such cases would be enough for my arguments to go through.

5   For these distinctions, see e.g. Harman (1986: 13–14). I here take explicit belief to be identical to occurrent belief, but Harman does not: he writes that "one believes something explicitly if one's belief in that thing involves an explicit mental representation whose content is the content of that belief" (1986: 13), and that "a belief is occurrent if it is either currently before one's consciousness or in some other way currently operative in guiding what one is thinking or doing" (1986: 14).

6   As Stevenson (2002: 106) writes, our concept of a belief "may well be vague in certain respects," and "may even be a family resemblance concept . . . with some of its extension left indeterminate by preceding usage." Stevenson (2002: 116–117, 120) distinguishes six different ways to use the term 'belief'; mine corresponds most closely to what he calls "linguistic reasoned beliefs."

7   In Streumer (2013a), I called these further conditions '(B1)' and '(B2)'. I here call them '(B3)' and '(B4)'.

8   For further discussion, see Streumer (2017: sect. 57). Partly because in Streumer (2013a) I did not make it clear that I used the term 'belief' to mean explicit belief, Bergamaschi Ganapini (2016) interprets (B3) as a claim about implicit beliefs. She argues that if we interpret (B3) in this way, my argument for the unbelievability of the error theory fails. As I said in Streumer (2016b), I agree. But I think that if we interpret (B3) as a claim about explicit beliefs, as I meant to do, my argument is sound.

9   I use the term 'evidence' to mean what Kelly (2006: sect. 3) calls 'indicator evidence': I use this term to mean a consideration that indicates that a belief is true, either by logically implying the content of this belief or by making it more likely that this content is true. By contrast, I use the term 'reason for belief' to mean a consideration that counts in favor of having this belief. We often take indicator

evidence to be a reason for a belief, but I think that reasons for belief are not the same thing as indicator evidence (see Streumer 2017: sect. 51).

10 As before, for further discussion, see Streumer (2017: sect. 57).

11 You may think that our inability to believe the error theory is a problem for this theory. But as I argue in Streumer (2013a; 2017: sect. 62), it is not. Just as a theory can be true if we do not believe it, a theory can also be true if we cannot believe it. Of course, if we cannot believe a theory, we cannot sincerely say that this theory is true. But that does not show that this theory is false.

12 I take a property to be descriptive if and only if it can be ascribed with a descriptive predicate. As I explain in Streumer (2017: sect. 45), this claim is compatible with different views about what *makes* a property normative or descriptive.

13 Alternatively, she may use the term 'evidence' to mean scientific evidence, and she may believe that there is non-scientific evidence for her belief that God exists (such as, perhaps, personal religious experience).

14 You may object that a reason to believe that p must be distinct from p itself. But on this conception of self-evidence, it is: what gives us sufficient reason to believe that p is not p itself, but our adequate understanding of p?

15 See Hyun and Sampson (2014: 634–635). They also list several other philosophical beliefs that they take to be counterexamples to (B4): the belief that my reason is reliable, the belief that I am not a brain in a vat, the belief that there are other minds, and the belief that my inductive reasoning is reliable. Lillehammer and Möller (2015: 455–456) give the last example as well. My claims about the belief that my senses are reliable also apply to these other purported counterexamples.

16 Hyun and Sampson (2014: 634) take this reason to be 'explanatory' rather than 'justificatory': they take it to be a consideration that explains Susan's belief but that does not count in favor of her belief. But if Susan *herself* does not take this reason to be justificatory, and if she also believes that there is no other justificatory reason for her belief that her senses are reliable, then I think she does not really believe that her senses are reliable.

17 Forcehimes and Talisse argue that if we believed the error theory, (B3) would also be false of us: they think that anyone "who fully believes entailments lack any reason-giving force can fail to believe what she fully believes to be entailed by one of her full beliefs" (2016: 854). But I do not think that (B3) is true of us because we believe that entailments have reason-giving force. Instead, I take (B3) to be true in virtue of what it is to believe that p, and in virtue of what it is to believe that p entails q.

18 Lillehammer and Möller (2015: 457–458) and Bergamaschi Ganapini (2016: 529–530) make similar claims. You may think Olson is himself a counterexample to (B4), since he may seem to believe the error theory while believing that there are no reasons for belief. But Olson actually does not believe that there are no reasons for belief: he thinks that judgments about reasons for belief can be judgments about instrumental reasons or judgments about what he calls 'the standard of being a responsible believer', in which case he thinks that his error theory does not apply to them. I discuss Olson's view in more detail in Streumer (2016a).

19 Even to young children, however, which beliefs they form on the basis of their perceptual inputs will not seem arbitrary. This may indicate that they do make implicit normative judgments when basing their beliefs on these inputs.

20 Owens (2000: 13) writes that "responsiveness to reasons does not require actual reflection on reasons: I can form a rational belief in p based on evidence e without forming either the belief that I have that evidence, or the belief that e suffices to justify p." But he also suggests that forming a belief that p based on evidence e *does* require *not* positively believing that I do *not* have this evidence. I think it similarly requires not positively believing that e is no reason to believe p.

21 You may object that if I do not believe that a consideration is a reason for a belief, I may nevertheless think that this consideration stands in some other positive normative relation to this belief: for example, that it justifies this belief, that it supports this belief, or that it makes it the case that I ought to have this belief. This shows that (4) should be generalized to: we can base a belief on a consideration *while* believing that this consideration *stands in no positive normative relation* to this belief. Since the error theory entails that these other positive normative relations also do not exist, this does not affect my argument.

22 A similar example is given by Evans (2013: 2946–2947).

23 For discussion of these views, see, for example, Korcz (2000; 2015) and Evans (2013). Korcz defends a combination of the first and the last views, and Evans defends a version of the second view.

24 Bergamaschi Ganapini (2016: 532) suggests that there are two related ways in which we can come to believe the error theory without believing that there is no reason to believe this theory: first, we may fail to realize "that the error theory entails that there are no reasons to believe it because [we have] never read or carefully analyzed this particular aspect of the error theory" (2016: 527), and second, we may "temporarily [ignore] that this is entailed by [our] belief about normative judgements in general" (2016: 532). My reply also applies to these suggestions.

25 They add that they "offer these considerations in a Moorean spirit, not as a decisive refutation of Streumer's view."

26 In Streumer (2017: sect. 63), I outline three further ways in which we can do this.

27 I give these arguments in Streumer (2008; 2011; 2013b; 2013c; 2017).

28 Of course, we can come to believe a moral error theory on the basis of, for example, Mackie's or Joyce's arguments for this theory. My claim is only that if we adequately understand my arguments for an error theory about all normative judgements, we cannot come to believe a moral error theory on the basis of these arguments.

29 As I said in Streumer (2013a: 203), you may think that if I really believed that these arguments are sound, I would believe that they *show* that the error theory is true. But I know that the claim that there are sound arguments that show that the error theory is true entails that the error theory is true. Given (B1), this means that I cannot believe that there are sound arguments that *show* that the error theory is true. But I can believe that there are sound arguments that *seem* to show this.

30 Since she is here objecting to the second way in which I think we can come close to believing the error theory, I have substituted the phrase "believe that there are sound arguments that together seem to show that *p*" for Bergamaschi Ganapini's phrases "come close to believing *p*" and "come close to having full confidence in the truth of a proposition." The latter phrase seems to equate coming close to believing the error theory with having a strong partial belief in the error theory. I do not think we can come close to believing the error theory in this way, since I think that any partial belief in the error theory must be very weak (see Streumer 2013a: 210; 2017: sect. 63).

31 At least, if we assume that we cannot believe (5) on the basis of (4) without thereby coming to believe that there is no reason to believe the error theory.

32 Have I reached this conclusion by clarifying the original argument or by slightly revising the argument? I think this does not matter. What matters is that the argument as I have formulated it here is sound, and that it makes the error theory more likely to be true in the ways I have outlined elsewhere (Streumer 2013a; Streumer 2017: sects. 72–74).

33 For helpful comments on earlier versions of this chapter, I am grateful to Hallvard Lillehammer, Diego Machuca, Mark van Roojen, and an audience at the Humboldt University of Berlin.

References

Bergamaschi Ganapini, M. 2016. "Why We Can Still Believe the Error Theory," *International Journal of Philosophical Studies* 24: 523–536.

Bortolotti, L. 2010. *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.

Cuneo, T. 2007. *The Normative Web*. Oxford: Oxford University Press.

Evans, I. 2013. "The Problem of the Basing Relation," *Synthese* 190: 2943–2957.

Forcehimes, A., & Talisse, R. 2016. "Belief and the Error Theory," *Ethical Theory and Moral Practice* 19: 849–856.

Harman, G. 1986. *Change in View*. Cambridge, MA: MIT Press.

Hyun, A., & Sampson, E. 2014. "On Believing the Error Theory," *Journal of Philosophy* 111: 631–640.

Kelly, T. 2006. "Evidence," *Stanford Encyclopedia of Philosophy*, Fall 2008 edition, http://plato.stanford.edu/archives/fall2008/entries/evidence/.

Korcz, K. 2000. "The Causal-Doxastic Theory of the Basing Relation," *Canadian Journal of Philosophy* 30: 525–550.

———. 2015. "The Epistemic Basing Relation," *Stanford Encyclopedia of Philosophy*, Fall 2015 edition, http://plato.stanford.edu/archives/fall2015/entries/basing-epistemic/.

Lillehammer, H., & Möller, N. 2015. "We Can Believe the Error Theory," *Ethical Theory and Moral Practice* 18: 453–459.

Olson, J. 2014. *Moral Error Theory*. Oxford: Oxford University Press.

Owens, D. 2000. *Reason without Freedom*. London: Routledge.

Stevenson, L. 2002. "Six Levels of Mentality," *Philosophical Explorations* 5: 105–124.

Streumer, B. 2008. "Are There Irreducibly Normative Properties?" *Australasian Journal of Philosophy* 86: 537–561.

———. 2011. "Are Normative Properties Descriptive Properties?" *Philosophical Studies* 154: 325–348.

———. 2013a. "Can We Believe the Error Theory?" *Journal of Philosophy* 110: 194–212.

———. 2013b. "Why There Really Are No Irreducibly Normative Properties." In D. Bakhurst, B. Hooker, & M. Little (eds.), *Thinking about Reasons: Essays in Honour of Jonathan Dancy*, 310–336. Oxford: Oxford University Press.

———. 2013c. "Do Normative Judgements Aim to Represent the World?" *Ratio* 26: 450–470.

———. 2016a. "Why Jonas Olson Cannot Believe the Error Theory Either," *Journal of Moral Philosophy* 13: 419–436.

———. 2016b. "No, We Cannot," *International Journal of Philosophical Studies* 24: 537–546.

———. 2017. *Unbelievable Errors*. Oxford: Oxford University Press.

5 Are There Substantive Moral Conceptual Truths?

David Copp

1. Introduction

Normative nonnaturalism is a kind of moral realism, at least in its familiar orthodox "robust" form. It agrees that there are true substantive moral claims. It might agree, for instance, that torture is wrong. It agrees that there are moral properties such as wrongness and that these properties are instantiated. It might agree that torture instantiates wrongness. It holds that moral properties are normative, and nonnaturalists typically would say that the normativity of these properties is a matter of their being a source of reasons. Importantly and characteristically, nonnaturalism rejects the thesis of normative naturalism that moral and other normative properties are natural ones. It holds that properties of these kinds are in a fundamentally different metaphysical category from garden-variety natural properties such as meteorological, psychological, or economic ones. David Enoch says, for example, that "in whatever sense there are physical facts [and properties], there are normative ones," yet "normative facts [and properties] are just too different from natural ones to be a subset thereof" (Enoch 2011: 4–5).

Orthodox nonnaturalism nevertheless faces a litany of familiar metaphysical and epistemological challenges (e.g., Enoch 2011: chs. 6–9). Recently, some philosophers have attempted to escape these challenges by developing forms of nonnaturalism that avoid any robust metaphysical commitments and that attempt in related ways to avoid the familiar epistemological challenges to orthodox nonnaturalism. Call views of this kind "Avant-Garde Nonnaturalism."

One such view agrees with orthodox nonnaturalism that there are moral properties such as wrongness, that these properties are instantiated, and that they are not natural ones, but it contends that these claims have no robust ontological implications. Following Derek Parfit's lead, we could call views of this kind "Non-Metaphysical Non-Naturalism" (2011: II, 486) or "Non-Realist Cognitivism" (2017: sect. 131).[1] I think we should set aside the question whether a view of this kind qualifies as a kind of moral or normative *realism*. This seems to me to be a terminological issue of little importance.

A second kind of avant-garde position contends that there is no need for the nonnaturalist to hold that moral *properties* are nonnatural; that non-naturalism requires only that there are irreducible nonnatural moral and normative *concepts* and true substantive moral propositions involving these concepts. In this way, it aims to avoid the metaphysical challenges to traditional nonnaturalism. And it aims to avoid the epistemological challenges by claiming that basic or fundamental substantive moral truths are *conceptual* truths, a category of truth of which we are familiar and that is, perhaps, not mysterious. In this way, it aims to do without the need to rest moral knowledge on the synthetic *a priori*. Call views of this kind "Conceptual Nonnaturalism." A bold version of Conceptual Nonnaturalism has been proposed by Cuneo and Shafer-Landau (2014). Their proposal, and the prospects for Conceptual Nonnaturalism more generally, are the topic of this essay.

Section 2 attempts to explain the motivation for Conceptual Nonnaturalism. It briefly sets out the familiar, standard, metaphysical and epistemological challenges to orthodox nonnaturalism and then explains how Conceptual Nonnaturalism seeks to avoid these challenges. Section 3 introduces Cuneo and Shafer-Landau's version of Conceptual Nonnaturalism, which they call "Minimal nonnaturalism" (2014: 403), as well as their central thesis, the "moral fixed points thesis," according to which propositions they call "the moral fixed points" are conceptual truths. Section 4 criticizes their theory of conceptual truth and argues that their theory of concepts faces challenges that are similar to those faced by orthodox nonnaturalism. It also considers an alternative account of the nature of concepts. Section 5 criticizes Cuneo and Shafer-Landau's claim that there are *substantive* moral claims that are conceptual truths. Section 6 considers the moral fixed points and the status they might plausibly have, given that they need to be qualified in ways I will discuss. The upshot is skeptical of the prospects for Conceptual Nonnaturalism.

2. Introducing Conceptual Nonnaturalism

Let us begin by asking what nonnaturalists mean to deny, in rejecting the thesis that normative properties are *natural* ones. On this point, if there is a consensus, it is perhaps that natural properties are properties of the kind studied in the sciences, including properties that may not yet be known but that will be studied in the best sciences of the future, and including properties that are reducible to properties of the kind just specified (see Parfit 2011: II, 305; Shafer-Landau 2003: 59). Now science is our best source of empirical or *a posteriori* knowledge of the world. Hence, this "science-based characterization" of the natural is a close relative of the suggestion that the natural properties are those of whose instantiation we can have only empirical or *a posteriori* knowledge, at least when it comes to *substantive* truths about their instantiation. This "empirical characterization" of the natural is more ecumenical than the science-based characterization since

it allows us to say that some prosaic and humdrum properties are natural ones even if they will never be studied in any science. On the empirical characterization, the properties of being a swimming pool and of having been born in September are natural properties, for example, but not, presumably, on the science-based characterization. In any case, I will use the empirical characterization in what follows.[2]

On the empirical characterization, the nonnaturalist is claiming, as against the naturalist, that moral and other normative properties are such that there can be *a priori* knowledge of *substantive* truths about their instantiation. The word "substantive" is doing important work here. A naturalist is not committed to denying that there can be conceptual truths about the instantiation of natural properties. It is a conceptual truth, for example, that deciduous trees are trees, and this does not mean that the property of being a tree is not a natural one. Similarly, a moral naturalist can concede that it is a conceptual truth that wrongness is a property of actions. Importantly, however, these truths about tree-hood and wrongness are not substantive.

Say that a proposition about the instantiation of a property is *substantive* if and only if a skeptic who denied the existence of the property would be committed thereby to denying the truth of that proposition.[3] To understand this, begin with a nonmoral example. It seems to me that stop signs are red only if there is such a thing as redness—only if there is a property of redness. Similarly, it seems to me, lying is wrong only if there is such a thing as wrongness—only if there is a property of wrongness. To be sure, nothing follows about the metaphysical nature of these properties. Acknowledging that stop signs are red does not commit us to a metaphysically robust account of properties. Nonetheless, in whatever sense it does commit us to there being a property of redness, the claim that lying is wrong likewise seems to commit us to there being a property of wrongness.[4] So on my proposal, the claim that lying is wrong is substantive. The claim that red things are red is not substantive, however, and, similarly, the claim that wrongful killings are wrong is not substantive. A skeptic about the existence of moral properties is not thereby committed to denying that wrongful killings are wrong![5] Substantive moral propositions, then, are those that a skeptic who denied the existence of moral properties would thereby be committed to denying. I shall say more about this idea in what follows.

What naturalism denies is that there is or can be *a priori* knowledge of *substantive* truths about the instantiation of a moral or normative property. Nonnaturalism holds, to the contrary, that there are such truths. Conceptual Nonnaturalism holds more specifically that there are substantive moral truths that are *conceptual* truths. It therefore has a neat response to the standard epistemological objections to orthodox nonnaturalism. On the empirical characterization, nonnaturalism might appear to be committed to the claim that there can be *a priori* knowledge of *substantive synthetic* truths about the instantiation of moral properties. It therefore has to face skepticism about the existence of substantive synthetic *a priori* knowledge.

Conceptual Nonnaturalism avoids such skepticism since its key claim is about the existence of *conceptual* truths, and, setting aside Quinean concerns, there is little reason to deny the existence of conceptual truths. The real issue is whether any conceptual truths are *substantive*.

Conceptual Nonnaturalism also has a neat response to the standard metaphysical objections to orthodox nonnaturalism, which concern the very idea of a nonnatural property. On the empirical characterization of the natural, the orthodox nonnaturalist is claiming, in effect, that reality includes at a fundamental level moral properties that are over and above those studied in the sciences and over and above those of whose instantiation we have empirical knowledge.[6] Accordingly, the nonnaturalist is committed, it seems, to admitting that moral properties are such that their instantiation can make no difference to our experience and that we cannot or at least do not interact with them causally. For if their instantiation might make a difference to our experience, or if we might interact with them causally, then science might take an interest in them and study them, and so, according to the empirical characterization, they would be natural properties. It appears, then, that orthodox nonnaturalism is committed to saying that there are in reality these special properties that can make no difference to our experience and with which we do not interact causally.

This might not seem to be a problem since mathematics is often taken to traffic in entities or properties that are over and above those amenable to empirical study. Yet, first, science is highly mathematical, so mathematical facts arguably do make a difference to our experience. Second, philosophers of mathematics struggle with the problem of how to understand its metaphysics (see Benacerraf 1965). It is not clear what mathematics "traffics in."

There are, I think, three metaphysical worries about the orthodox view, if I am correct to have characterized it as I have. First, moral facts are taken to be of great importance to us in thinking about how to live our lives. It can seem bizarre to suppose that properties that can make no difference to our experience and with which we do not interact causally can be of such great importance. Second, it can seem bizarre to suppose that the property wrongness has the same kind of status in the fundamental nature of reality as do the fundamental properties of physics. And finally, third, is an objection that is partly epistemological. It might seem that there cannot be good grounds to suppose there are nonnatural properties—properties that can make no difference to our experience and with which we do not interact causally.

Conceptual Nonnaturalism has a neat response to these worries. For, as I understand the view, it does not postulate the existence of *nonnatural properties*. It only postulates the existence of *nonnatural concepts*. And it claims that we can have *a priori* knowledge of substantive conceptual truths about how to live our lives. It makes no claim about our ability to gain *a priori* access to fundamental truths about the *properties* found in reality. As such, Conceptual Nonnaturalism is compatible with metaphysical naturalism, according to which the natural exhausts reality. Some versions of

Conceptual Nonnaturalism might also be compatible with orthodox nonnaturalism, for, as I understand it, although Conceptual Nonnaturalism does not claim that there are nonnatural moral properties, it need not *deny* that there are.[7]

Clearly, then, Conceptual Nonnaturalism is a promising approach to defending nonnaturalism against the standard objections to orthodox nonnaturalism. If there genuinely is room in logical space for the thesis that there are *substantive conceptual* moral truths, Conceptual Nonnaturalism would appear to be a viable new approach to understanding the relation between the moral truths and natural facts about the empirical world.

3. Cuneo and Shafer-Landau's Minimal Nonnaturalism

Cuneo and Shafer-Landau have blazed a route into this under-explored region of logical space. They take the "Core Claim" that characterizes nonnaturalism to be the thesis that there are nonnatural moral truths (2014: 401). And they explain that a nonnatural moral truth is a proposition that is not identical with any naturalistic proposition, nor is it "made true exclusively by [any] natural fact."[8] They then distinguish between the following views (2014: 403):

> *Minimal Nonnaturalism*: there are nonnatural moral truths, but there are no nonnatural moral properties or facts. All moral properties and facts are natural.
> *Robust Nonnaturalism*: there are both nonnatural moral truths and nonnatural moral properties and facts.

Robust Nonnaturalism is the view I have been calling orthodox nonnaturalism. Cuneo and Shafer-Landau say that they favor a position that is close to the robust view. Nevertheless, they say, both Minimal and Robust Nonnaturalism are compatible with the main thesis they aim to defend, which is that "there is a range of moral propositions, the moral fixed points," which "are constituted by nonnatural moral concepts," and which are not "made true exclusively by natural facts." Rather, they say, these propositions "are true in virtue of the nature of the nonnatural moral concepts that constitute them" (2014: 403). Call this the "moral fixed points thesis."

Cuneo and Shafer-Landau give ten examples of moral fixed points, of which the following four seem to be representative (2014: 405):

> MFP1: It is *pro tanto* wrong to engage in the recreational slaughter of a fellow person.
> MFP2: It is *pro tanto* wrong to humiliate others simply for pleasure.
> MFP3: There is some moral reason to offer aid to those in distress, if such aid is very easily given and comes at very little expense.

> MFP4: It is *pro tanto* wrong to satisfy a mild desire if this requires killing many innocent people.

According to Cuneo and Shafer-Landau, the fixed point propositions set conceptual limits to what can count as a moral system for "beings like us in a world such as ours" (2014: 404). The propositions should be qualified accordingly. They need to be qualified in order to side-step objections that rest on bizarre counter-examples (2014: 405, n. 15). Hence, for example, MFP2 should strictly speaking read: *For beings like us in worlds like ours*, it is *pro tanto* wrong to humiliate others simply for pleasure (2014: 405). In what follows, I will sometimes focus on MFP2, since it is the briefest of the lot and since I do not think anything turns on the differences among these propositions.

One might wonder why the fixed point propositions should be classified as *nonnatural* truths. Cuneo and Shafer-Landau take it that propositions are at least partially constituted by the concepts one would need to have in order to grasp them (2014: 410). And they observe that each fixed point proposition is partly constituted by a moral concept. MFP2 is partly constituted by the concept of wrongness, for example. They further claim that moral concepts and normative concepts more generally are not, and are not reducible to or analyzable in terms of, naturalistic concepts (2014: 402–403). Many naturalists would agree with this, of course (e.g., Railton 2003). It is arguable, for example, that the concept of wrongness is the concept, roughly, of a property actions can have of being in violation of an authoritative moral standard.[9] Naturalists can agree that this concept is different from any naturalistic concept and that it is not reducible to or analyzable in terms of naturalistic concepts. For the central concern of naturalists is to claim that moral concepts refer *to natural properties* and that *substantive* moral truths are made true by *natural facts*.[10] It appears, then, that naturalists can agree with Cuneo and Shafer-Landau that the fixed point propositions are not, and are not reducible to or analyzable in terms of, naturalistic propositions.[11] I will return to this point briefly in the next section.

It is important to understand that naturalism is compatible with recognizing that some moral propositions are conceptual truths. An example is the proposition that murder is wrong. Even more trivial is the proposition that wrongful killing is wrong. A naturalist would insist that such propositions are not substantive, but she need not deny that such propositions are true in virtue of the relations between the "concepts that constitute them" (2014: 403). It therefore appears that naturalists can accept Cuneo and Shafer-Landau's moral fixed points thesis.

Furthermore, given what I have just explained, naturalists can also accept Cuneo and Shafer-Landau's "Core Claim." This is the thesis that there are nonnatural moral truths, where, Cuneo and Shafer-Landau explain, a nonnatural moral truth is a proposition that is not identical with any naturalistic proposition nor is it "made true exclusively by [any] natural fact" (2014: 401).

We have just seen that a naturalist can agree that a moral proposition is not identical to any naturalistic proposition and she can agree that moral conceptual truths are not made true exclusively by natural facts but that they are, rather, true in virtue of relations between the "concepts that constitute them." So a naturalist can accept the Core Claim. And she can also accept "Minimal nonnaturalism," the thesis that "there are nonnatural moral truths, but there are no nonnatural moral properties or facts," and "all moral properties and facts are natural." We have seen that a naturalist can accept that there are nonnatural moral truths, and she can obviously agree that all moral properties and facts are natural.

There is, then, the question why Cuneo and Shafer-Landau view their Core Claim as characterizing *nonnaturalism* and why they view so-called Minimal Nonnaturalism as a kind of *nonnaturalism*. I suspect that they simply overlooked the fact that a naturalist can allow that there are trivial, tautological nonnatural moral truths, such as, say, the truth that wrongful killing is wrong. If we amend the Core Claim and the thesis of Minimal non-naturalism slightly, so that they claim that there are *substantive* nonnatural moral truths, then I believe, naturalists would deny both theses. A naturalist would want to claim that any *substantive* moral truth is true in virtue of the obtaining of relevant natural facts and that the moral concepts refer to natural properties. As against this, Cuneo and Shafer-Landau (2014: 408, 410) claim that the fixed point propositions are *substantive* truths that are true in virtue of the nature of the concepts that constitute them. I believe, then, that Cuneo and Shafer-Landau intend the Core Claim and Minimal nonnaturalism to entail that there are substantive nonnatural moral truths. This is why they view these claims as staking out a version of nonnaturalism.

Now, if we combine the moral fixed points thesis with the Core Claim, amended as I have just suggested, we can derive the following thesis:

> *The Embellished Core Claim*: There are [substantive] nonnatural moral truths. These truths include the moral fixed points, which are a species of conceptual truth, as they are propositions that are true in virtue of the essences of their constituent concepts.
>
> (Cuneo & Shafer-Landau 2014: 411–412)

Naturalists would deny this claim because it entails that there are substantive conceptual moral truths.[12]

The main reason for accepting the Embellished Core Claim, say Cuneo and Shafer-Landau (2014: 422), is its explanatory power. First, the idea that the fixed point propositions are *conceptual* truths explains why they seem necessary and incontestable, why they might seem self-evident, and why acceptance of them seems to be criterial of whether a person has a basic moral competence (2014: 406–408). Second, as I explained before, the Embellished Core Claim seems to dodge the familiar objections to ortho-dox nonnaturalism (2014: 422–431). "For conceptual truths do not require

facts that correspond to them, worldly truth-makers that render them true" (2014: 411). Moreover, "whatever account of conceptual knowledge is best," nonnaturalists can use it explaining "how knowledge of the moral fixed points is possible" (2014: 437). The issue remains, however, as to why we should believe that the fixed point propositions are substantive truths. And here, I think, Cuneo and Shafer-Landau would argue by ostension. MFP2 certainly seems to be substantive, for example. I will return to this issue, but I need first to discuss Cuneo and Shafer-Landau's theory of concepts and conceptual truth.

4.   Concepts and Conceptual Truths

Conceptual Nonnaturalism seeks to avoid familiar objections to orthodox nonnaturalism since it does not postulate nonnatural moral properties or facts. Indeed, Cuneo and Shafer-Landau's Minimal naturalism holds that moral properties and facts are natural ones. But Conceptual Nonnaturalism does postulate nonnatural concepts and it contends that there are *substantive* conceptual moral truths. Its plausibility therefore depends on the view we take of the nature of concepts, propositions, and conceptual truths. Unfortunately, as we will see, on Cuneo and Shafer-Landau's theory of concepts, concepts are metaphysically akin to properties as a Platonist would understand them. Because of this, I contend, their theory of concepts and conceptual truth faces epistemic and metaphysical challenges that are akin to, even if somewhat different from, those faced by orthodox nonnaturalist theories. If Conceptual Nonnaturalism is to achieve its goal of avoiding these objections, I believe it needs to be allied with a different theory of concepts than Cuneo and Shafer-Landau's theory. Let me explain.

Cuneo and Shafer-Landau's position rests on a view of concepts and conceptual truths that they call the "traditional view" (2014: 409). I will call it the "Platonistic view" for reasons that will become evident. First, on their view, concepts are "abstract," "extra-mental," "mind-independent" ways of thinking about objects and properties, the existence of which does not depend on our using them in either thought or language (2014: 409). Although, they say, concepts are "typically" identified with the meanings of words, they do not commit to this. They hold that "not all concepts are expressed linguistically. Some we have yet to discover; others may permanently elude us, owing, perhaps, to their complexity" (2014: 411). Second, in their view, concepts are among the constituents of propositions, which are the objects of belief (2014: 409–410). Third, unlike properties, concepts are "referential devices," and in standard cases, "concepts necessarily determine the range of entities that they are about" (2014: 410). The concept of wrongness, for example, is necessarily about wrongness (2014: 410). Fourth, according to Cuneo and Shafer-Landau, concepts have essences. The concept of wrongness is such that "it belongs to the essence of the concept that it applies to exactly those things that are wrong (if any such things there be)" (2014: 410).

This is the key to their theory of conceptual truth. In paradigm cases, they say, conceptual truths are true in virtue of the essences of their constituent concepts. The proposition <that *x* is F> is a conceptual truth if and only if it belongs to the essence of the concept of F, or to the essence of the concept of *x*, or both, that, necessarily, anything that satisfies the concept of *x* also satisfies the concept of F (2014: 410). Suppose that the essence of the concept of wrongness is such that, necessarily, anything that satisfies the concept of being a humiliation of a person simply for pleasure also satisfies the concept of wrongness. If so, it is a conceptual truth that humiliating a person just for pleasure is wrong. The truth of this proposition is due to the relation among its constituent concepts and does not depend solely (if at all) on the existence of any correlative worldly facts (2014: 410).

Let me emphasize that, on Cuneo and Shafer-Landau's view, concepts are abstract, extra-mental, and mind-independent, and their existence does not depend on our using them in either thought or language (2014: 409). It follows that the concept of wrongness would exist even if no one had ever noticed it. Concepts are another layer of abstract entity alongside properties. This aspect of their view gives rise to problems.

It is unclear, for instance, how we are supposed to grasp concepts, or have access to them, and why we should suppose any such things exist, on Cuneo and Shafer-Landau's Platonistic account of what they are. On a familiar account of the matter, we learn what wrongness is, as children, as a result of being taught what kinds of actions to avoid. A child might be told, "Don't cheat! Cheating is wrong!" Through teaching of this kind, she might come to have an inventory of paradigm examples of wrongful behavior and of how wrongful behavior is regarded. She learns that wrongful behavior is regarded as behavior that violates an important moral norm, and behavior for which she will be blamed unless she has an excuse. As a result of some such process—please excuse the sketchiness of the account—we come to understand the meaning of "wrong" and we come to know what wrongness is. We come to have thoughts about wrongdoing. Now, it seems to me, it is implausible that this process of moral learning involves our grasping a "concept" of wrongness, understood as a special kind of Platonistic entity. On the Platonistic view, the concept of wrongness is a kind of intermediary such that our having the concept accounts for our ability to have, for instance, thoughts about wrongdoing. If we needed to postulate such an intermediary, wouldn't we need in turn to postulate a second-order intermediary? Wouldn't we need to postulate a second-order concept, the concept of the concept of wrongness, such that our grasping this concept accounts for our ability to grasp the first-order concept of wrongness just as, on Cuneo and Shafer-Landau's view, our grasping the first-order concept accounts for our ability to grasp the property of wrongness? It is not clear why we need to suppose that coming to have beliefs that are about properties requires grasping corresponding concepts, understood as Cuneo and Shafer-Landau do, as abstract entities that exist alongside properties.

As Alan Gibbard (2003: 25) has argued, we need to suppose that we have concepts of things, and that we can have different concepts of one and the same thing, in order to explain the possibility of coherently accepting and denying propositions that have the same truth conditions. A person can coherently believe that the glass is full of water, but deny that it is full of H_2O molecules, even though—I assume—to be water is to be composed (perhaps mainly) of H_2O molecules. To explain such cases, we postulate two concepts of water, the ordinary and the chemical concepts. But nothing in this line of reasoning argues for a Platonistic theory of concepts as opposed to alternatives, such as a functional account.

On a functional account, *having a concept* is identified with a functional state within an overall functional account of intentional psychological states, such as belief. On such an account, roughly speaking, when the functional account of belief distinguishes between beliefs that have the same truth conditions, such as the belief that the glass is full of water and the belief that it is full of H_2O molecules, the functional account of concepts distinguishes corresponding concepts, such as the ordinary and the chemical concepts of water. Parenthetically, note that a functional account helps to clarify how it is that a moral naturalist can agree that there are nonnatural concepts. For even if a naturalist holds that the property wrongness is identical, say, to the property of failing to maximize the general welfare, she can accept that the belief that lying is wrong is distinct from the belief that lying fails to maximize the general welfare. And, accordingly, she can agree that the moral concept of wrongness is distinct from the naturalistic concept of failing to maximize the general welfare. In any event, the important point for present purposes is that a functional account of what it is to have a concept does not postulate Platonistic concepts as a special kind of abstract entity. It rather postulates a certain kind of representational state of mind, that of "having a concept" or "having a way of thinking" of something. Call this the "functional theory of concepts."

To be sure, I have done no more than to sketch a functional theory. Indeed, what I have said is less than a sketch. It would take me too far afield to try to develop the idea. For present purposes, the most important feature that is missing from what I have said is a story about conceptual truth. It is not clear that a functional theory of concepts is positioned to yield an account of conceptual truth that could do the work, in a version of Conceptual Nonnaturalism, that is done by Cuneo and Shafer-Landau's Platonistic theory. Conceptual Nonnaturalism rests on the idea that there can be *substantive* conceptual truths, for, as we saw, this idea is what marks it as a form of *nonnaturalism*. It is not clear that the functional theory is suited to support this idea. It may be that the claim that there can be substantive conceptual moral truths depends on a Platonistic theory of concepts much like Cuneo and Shafer-Landau's view. I therefore return to an assessment of their view.

One mysterious feature of their view is the idea that, at least in central cases, concepts "necessarily determine the range of entities that they are

about" (2014: 410). They add: "It belongs to the essence of the concept [of wrongness] that it applies to exactly those things that are wrong" (if any are). Consider the concept of bachelor and suppose that Alan is a bachelor. The idea is not, of course, that it is of the essence of the concept of bachelor that it applies to Alan. Rather, the idea seems to be that it belongs to the essence of the concept that, if anyone is a bachelor, then the concept bachelor applies to that person. It is a conceptual truth that the concept bachelor applies to Alan if he is a bachelor. Now this claim is meant to apply to all concepts (except indexical and demonstrative concepts) (2014: 410, n. 25). Hence, the claim seems to be that, for any property P, if it is represented by a concept, CP, then it is of the essence of CP that if anything has P then CP applies to it as well.[13] In the case of any property that is represented by a concept, the concept has an essence that determines that its extension matches the extension of the property.

On Cuneo and Shafer-Landau's view, then, there are two parts to Plato's heaven. One part contains properties. Another part is reserved for concepts, where the concepts represent the properties (2014: 411).[14] Plato's heaven for concepts would seem to be full of a multiplicity of concepts, perhaps one for each property,[15] including many concepts that no one has ever grasped, and perhaps even many that do not correspond to a property. And Plato's heaven for properties would seem to be full of a multiplicity of properties, including many that are never to be instantiated in the actual world. Cuneo and Shafer-Landau hold that, necessarily, each property that is represented by a concept is such that the nature of this concept determines that it applies to anything that has the property. This would be a necessary connection between "distinct existences" and it would seem to need explanation.

As we have seen, there can be different concepts that refer to one and the same property, such as the ordinary and the chemical concepts of the property of being water, both of which refer to this property. To the extent that I have a grip on the idea that a concept can have an essence, it seems clear that the chemical concept of this property and the ordinary concept have different essences, even if the essence of each of them determines that its extension matches the extension of the property. So, it seems, different concepts of a thing can have different essences. Given this, the question arises whether there might be different concepts of moral properties such as wrongness. I don't see any machinery in the Platonistic view to rule this out, nor to rule out that these concepts might have significantly different essences. This would, however, be problematic for Cuneo and Shafer-Landau's defense of the fixed points thesis.

First, if there can be different concepts that represent wrongness, and if these concepts can have different essences, then the "conceptual truths" about wrongness would or might vary from concept to concept. The wrongness-1 concept might be such that it is of its essence that it is wrong to humiliate others just for fun. Relative to this concept, it would be a conceptual truth that it is wrong to humiliate others just for fun. The wrongness-2

concept might have no such essential property. Relative to that concept, it would not be conceptually guaranteed that it is wrong to humiliate others just for fun. And the wrongness-3 concept might be such that it is of its essence that it is not morally wrong, but is instead good and admirable, to humiliate others just for fun. Hence it could turn out that even if all of Cuneo and Shafer-Landau's fixed points are conceptual truths with respect to one of the wrongness concepts, some might not be conceptual truths relative to other wrongness concepts, and some might be conceptual falsehoods relative to still other wrongness concepts.[16]

Cuneo and Shafer-Landau would deny that there could be a wrongness-3 concept that has a different extension than do the wrongness-1 and wrongness-2 concepts. For they hold that any concept that represents wrongness must have an essence that determines that its extension matches the extension of the property. They would say that if wrongness-3 has a different extension than wrongness-1 and wrongness-2, these cannot all be concepts of wrongness. But I don't see any argument to show that every concept of a property must have an essence that guarantees that it has the same extension as the property.

Cuneo and Shafer-Landau might respond that the moral fixed points set the boundaries of morality such that any *moral* system must accept (enough of) the fixed points (2014: 401, n. 6; 406). This is presumably intended as a conceptual point. It would follow that if enough of the fixed points are not true with respect to a concept, it is not a concept of *moral* wrongness. The problem is that, corresponding to the multiple concepts of wrongness, there could be multiple concepts of morality, each of which certifies almost all of the fixed points. Indeed, it strikes me as plausible that there are different concepts of morality. The concept of morality-1 might be such as to necessarily certify the fixed points guaranteed by the essence of the concept of wrongness-1. The concept of morality-2 might be such as to necessarily certify the fixed points guaranteed by the essence of the concept of wrongness-2. And so on.

Second, and worse, since the propositions that are the objects of our beliefs are constituted at least in part by relevant concepts, on Cuneo and Shafer-Landau's view, if it is indeterminate which of the wrongness concepts we have, it will be indeterminate which wrongness beliefs we have. People whom we intuitively take to disagree might not disagree on Cuneo and Shafer-Landau's view, since they might have beliefs that have different wrongness concepts as constituents. On their view, what determines which wrongness concept is a constituent of my beliefs about wrong action?

Cuneo and Shafer-Landau might respond that there is no room for the existence of different concepts of wrongness. For there is no possibility of a person's coherently believing that something is wrong-1 but not wrong-2 in the way that there is room for a person coherently to believe the glass is full of water but not of H_2O molecules. We have no words to distinguish these beliefs and we cannot otherwise distinguish them. Unfortunately, however,

on Cuneo and Shafer-Landau's Platonistic view, concepts are mind-independent. The question whether there is more than one wrongness concept does not depend on whether we have detected them, nor does it depend on facts about our beliefs. So, on their view, it seems that the belief you express by saying "Humiliating others for fun is wrong" might have as a constituent the wrongness-1 concept, and the belief I express using these words might have as a constituent the wrongness-2 concept. You might *assert* that it is a conceptual truth that humiliating others for fun "is wrong," and the belief you thereby express might not in fact be inconsistent with the belief I express by *denying* that it is a conceptual truth that humiliating others for fun "is wrong."

One might propose, as a friendly amendment to Cuneo and Shafer-Landau's view, that we should take concepts to be meanings. On the resulting amended view, the meaning of "wrong" would be "the" concept of wrongness. But if there can be more than one concept of wrongness, it might be indeterminate which of these concepts is the meaning of "wrong." It could be, on this view, that "wrong" has different meanings depending on which wrongness concept a given speaker associates it with. Or if we want to insist that "wrong" is univocal, the view would face the problem of explaining which wrongness concept is its meaning.

Something clearly has gone wrong with Cuneo and Shafer-Landau's theory of concepts and conceptual truth. Let me conclude this section of the essay by summarizing the main point.

Conceptual Nonnaturalism appears to avoid familiar objections to orthodox nonnaturalism since it does not postulate nonnatural properties. It instead postulates nonnatural moral concepts and, to distinguish itself from naturalism, it proposes that there are substantive conceptual moral truths. Cuneo and Shafer-Landau propose a Platonistic account of concepts and combine it with their "Minimal nonnaturalism" and their "fixed points thesis" to yield a version of Conceptual Nonnaturalism. But there are problems.

For one thing, there seem to be similar problems explaining our knowledge or awareness of Platonistic concepts as there are explaining our knowledge or awareness of nonnatural properties of the kind that are postulated by orthodox nonnaturalism. So it is not clear that the combination of Conceptual Nonnaturalism with Cuneo and Shafer-Landau's Platonistic view of concepts is preferable on metaphysical or epistemological grounds to orthodox nonnaturalism. It is true that Cuneo and Shafer-Landau's view purports to establish the fixed point propositions by construing them as conceptual truths underwritten by the essences of the relevant nonnatural moral concepts. But an orthodox nonnaturalist could instead argue that these propositions are metaphysically necessary truths underwritten by the essences of relevant nonnatural moral properties. This would mean abandoning the strategy of Conceptual Nonnaturalism, yet I don't see that the one approach to underwriting the necessity of the fixed point propositions

would be any more plausible or any less problematic than the other. For these reasons, it is not clear to me that, when it is combined with the Platonistic theory of concepts, Conceptual Nonnaturalism is any less metaphysically and epistemologically fraught than orthodox nonnaturalism.

If I am correct, the Platonistic account of concepts faces a series of additional problems. Given this, it seems to me, a viable form of Conceptual Nonnaturalism would need to offer a different theory of concepts and conceptual truth. I suggested that a functional theory might be viable. The trouble is that Conceptual Nonnaturalism rests on the idea that there can be substantive conceptual truths, and it is not clear that a functional theory of concepts would substantiate this idea. It may be that the claim that there can be such truths depends on a Platonistic theory of concepts much like Cuneo and Shafer-Landau's view.

5.   Are There Substantive Moral Conceptual Truths?

As we have seen, a naturalist can agree that there are moral conceptual truths, such as the proposition that wrongful killing is wrong. It is even open to a naturalist to agree that Cuneo and Shafer-Landau's fixed point propositions are conceptual truths. More generally, as we saw, a naturalist is not committed to denying that there can be conceptual truths about the instantiation of natural properties. It is a conceptual truth that deciduous trees are trees, and this does not mean that the property of being a tree is not natural. It is important, however, that these truths about wrongness and tree-hood are *non-substantive*. On the empirical conception of naturalism, a naturalist will say that any *substantive* truth about the instantiation of moral properties is empirical. The Conceptual Nonnaturalist would want to claim to the contrary that there are substantive, non-trivial moral truths that are conceptual truths. This is certainly something Cuneo and Shafer-Landau (2014: 400) want to claim.

The word "substantive" is doing important work, but what is meant? On the account I proposed in Section 2, a proposition about the instantiation of a property is substantive if and only if a skeptic who denied the existence of the property would be committed thereby to denying the truth of the proposition. If, for example, there is no property of wrongness, then the proposition that lying is wrong is not true.[17] Lying could not be wrong, one might say, if there is no such thing as wrongness. Now there actually are skeptics about the existence of moral properties. J. L. Mackie defended an error theory according to which, on one interpretation, there are no moral properties (1977).[18] Let me therefore refer to the view that there are no moral properties as the "error theory," leaving aside issues about how best to interpret Mackie's view. The question, then, is whether there are any *conceptual* moral truths that a proponent of the error theory would be committed to denying. These would be substantive.

Notice that if there is a conceptual truth that a proponent of the error theory would be committed to denying, then the error theory must entail that this conceptual *truth* is *not* true, and this means that the error theory must be false.[19] Moreover, the theory must be *conceptually* false, if its falsity is entailed by a conceptual truth. Its conceptual falsity would be entailed, for instance, by MFP2, if MFP2 were a substantive conceptual truth that a proponent of the error theory would be committed to denying.

It follows that we can investigate whether there are substantive conceptual moral truths by investigating whether the error theory is conceptually false. If the theory is *not* conceptually false, then there are no substantive moral conceptual truths. Now, it seems to me highly implausible that it is a conceptual truth that the property wrongness exists. If so, it is highly implausible that the error theory is conceptually false. And if so, then it is highly implausible that there are substantive conceptual moral truths—conceptual truths that one who embraced the error theory would thereby be committed to denying.[20]

Cuneo and Shafer-Landau anticipate this issue and respond that their view does not imply that it is a conceptual truth that the property wrongness exists. They say their view only implies, for instance, that

> it is a conceptual truth that, in worlds like ours and for creatures such as us, the concept 'being wrong' is such that, if anything satisfies a concept such as 'recreational slaughter,' then it also satisfies the concept 'being wrong.'
>
> (2014: 414, n. 33)

They point out that it does not follow from this claim about concepts that the property wrongness exists, "let alone that such a property exists as a matter of conceptual necessity" (2014: 414, n. 33). Nevertheless, it does follow, on their view, that recreational slaughter is wrong. According to their fixed points thesis, claims such as MFP1 and MFP2 are true, and these claims are about the wrongness of recreational slaughter and humiliating people, respectively, not about relations among concepts. The relevant issue is whether claims such as MFP1 and MFP2 entail (or perhaps presuppose) that the property of wrongness exists. If they do, then, I say, Cuneo and Shafer-Landau's claim that MFP1 and MFP2 are *conceptual truths* entails that it is a *conceptual truth* that wrongness exists.

One might try to show in a different way that the moral fixed points do not entail (or presuppose) that there is a property of wrongness. Consider, for example, the claim that wrongful killing is wrong. This claim does not entail or presuppose that there is a property of wrongness. An error theorist could consistently allow that any acts of wrongful killing would be wrong. Similarly, one might think, MFP1 does not entail or presuppose that there is a property of wrongness. It is equivalent, roughly, to the proposition that any acts of recreational slaughter would be wrong. And one might suggest that the concept of recreational slaughter is similar to the concept of

a wrongful killing in that it is partly constituted by the concept of wrongness such that it is *tautological* that acts of recreational slaughter would be wrong. This seems clearly to be a mistake, however, and it is a mistake that Cuneo and Shafer-Landau do not make.[21] A skeptic who denies that there is such a thing as wrongness is committed thereby to denying that there could be acts of wrongful killing. But she is not committed thereby to denying that there could be acts of recreational slaughter. She is committed, rather, to denying that acts of recreational slaughter would be wrong. For if there is no such thing as wrongness, it is not the case that recreational slaughter is wrong. The proposition that recreational slaughter is wrong *does* seem to entail (or presuppose) that there is such a thing as wrongness.[22]

The next issue is whether the thesis that MFP1 is a *conceptual truth* entails that it is a *conceptual truth* that wrongness exists. Here I assume that if <p> entails or presupposes <q> as a matter of conceptual truth, then <q> is a conceptual truth if <p> is a conceptual truth. I have just provided a conceptual argument that MFP1, the proposition that recreational slaughter is wrong, entails (or presupposes) that there is such a thing as wrongness. And similar arguments could be provided for the other fixed point propositions. So it seems to me that Cuneo and Shafer-Landau's thesis that the moral fixed point propositions are conceptual truths entails (or presupposes) that it is a conceptual truth that wrongness exists.

Recall that, by the "error theory," I here mean the thesis that there are no moral properties. Given this, the fixed points thesis directly entails that the error theory is conceptually false, for it says that there are conceptually true first-order moral claims that predicate wrongness of things. An example is MFP2: it is *pro tanto* wrong to humiliate others simply for pleasure. It seems to me that if MFP2 is a conceptual truth, as Cuneo and Shafer-Landau claim, then it is a conceptual truth that there is a property of wrongness.[23] Hence, it seems to me, Cuneo and Shafer-Landau are committed to the thesis that the error theory is conceptually false.

If this is correct, we can draw two conclusions. First, if the moral fixed point propositions are substantive propositions, a proponent of the error theory is committed to denying them. Second, if they are substantive, then the plausibility that they are conceptual truths depends on the plausibility of the thesis that it is a conceptual truth that there is a property of wrongness.

I think that this is highly implausible. Cuneo and Shafer-Landau ought to agree, for on their view, a proposition of the form <that *x* is F> is a conceptual truth just in case it belongs to the essence of the concept of F, or to the essence of the concept of *x*, or both, that, necessarily, anything that satisfies the concept of *x* also satisfies the concept of F (2014: 410, with n. 27). But it is highly implausible that the essence of the concept of *existence* is such that anything that satisfies the concept of being the property of wrongness also satisfies the concept of existence. And it is highly implausible that the essence of the concept of being the property of *wrongness*, or the essences of these concepts taken together, is such that anything that satisfies the concept

of being the property of wrongness also satisfies the concept of existence. Conceptual truths are underwritten by relations among concepts. But the existence of a property is not a matter merely of relations among concepts.

I conclude on this basis that it is implausible that there are substantive moral propositions that are conceptual truths. A substantive moral truth would be one that an error theorist would be committed to denying. Its falsity would be entailed or presupposed by the error theory. If such a claim were a conceptual truth, the error theory would be conceptually false. Yet it is not plausible that the error theory is conceptually false, so it is not plausible that there are substantive conceptual moral truths.

6.   The Moral Fixed Points

Despite the conclusion of the previous section, we need to look more closely at Cuneo and Shafer-Landau's moral fixed point propositions. The argument of the previous section has shown, if I am correct, that these propositions plausibly are *substantive* moral claims, but that since it is not plausible that the error theory is conceptually false, it is not plausible that they are *conceptual* truths. To support this conclusion, however, it is important to investigate these propositions more directly. I contend that even if we set aside the arguments we have explored so far, it is implausible that the fixed point propositions are conceptual truths.

Cuneo and Shafer-Landau make a number of claims in arguing for their thesis that the fixed point propositions are conceptual truths (2014: 407–408). They claim that these propositions are necessarily true if they are true at all. They are good candidates for being knowable *a priori*. They enjoy "framework status," fixing the boundary of morality for beings like us in worlds like ours. Most important, denial of these propositions "would tend to evoke bafflement among those competent with their constituent concepts" (2014: 408). But a moral naturalist can account for these claims, it seems to me. As I have contended elsewhere (Copp 2007c; 2007d), she can argue that the fixed point propositions are "moral necessities" and are likely to be "self-evident" in a sense that a naturalist can accept. I will not repeat these arguments here. In any case, Cuneo and Shafer-Landau concede that their claims do not show decisively that the fixed point propositions are conceptual truths.

The most important and obvious problem with the idea that the fixed point propositions are conceptual truths is that, as Cuneo and Shafer-Landau explain, when these propositions are spelled out in detail, they are qualified so as to contain reference to "worlds like ours" and "beings like us." Cuneo and Shafer-Landau (2014: 405) explain that their interest is in what moral systems must be like for people like us in situations such as we face, so they have no need to defend the stronger unqualified propositions. Furthermore, incorporating appropriate qualifications, they say, enables them to avoid objections about "bizarre possibilities" such as "humiliation friendly

worlds" in which humiliating people for pleasure is not wrong because, say, it helps people to strengthen their moral character (2014: 405, n. 15). For these reasons, strictly speaking, the fixed point propositions should all be qualified (2014: 405). Strictly speaking, MFP2 is the proposition that [*for beings like us in worlds like ours*, it is *pro tanto* wrong to humiliate others simply for pleasure]. Or as I will say, to avoid irrelevant complications, it is the proposition that [for beings like us in worlds like ours, it is *pro tanto* wrong to humiliate *a person* simply for pleasure].

There are two ways to interpret Cuneo and Shafer-Landau's proposed amendments to the fixed point propositions. The first interpretation is suggested by their wording, so it is the one we ought to favor, other things being equal. On this interpretation, the fixed point propositions need to be stated in a way that includes the indexical expressions "beings like us" and "worlds like ours" or, perhaps better, "beings like us, as we actually are" and "worlds such as ours, as it actually is." On this reading, MFP2 would be the proposition that [it is wrong for beings like us as we actually are, in worlds such as this world actually is, to humiliate people just for pleasure].[24] Since this proposition would be partly constituted by indexical concepts, in Cuneo and Shafer-Landau's view, I will call this the "indexical reading." The fixed points thesis would claim that MFP2 is a conceptual truth on the indexical reading. On a second interpretation, the fixed point propositions instead include descriptions of kinds of worlds and kinds of agents where, in fact, the described worlds are relevantly like our world and the described agents are relevantly like us. On this reading, MFP2 would be the proposition that [it is wrong for beings with properties A, B, and C in worlds with properties X, Y, and Z to humiliate people just for pleasure]—where, in fact, the actual world is one of the ABC and XYZ worlds. Call this the "descriptions reading." The fixed points thesis would then claim that MFP2 is a conceptual truth on the descriptions reading. There are problems for Cuneo and Shafer-Landau on both readings.

On the descriptions reading, even if MFP2 is a conceptual truth, it does not follow that it is a conceptual truth that we would actually be wrong in the actual world to humiliate others just for pleasure, since it is not a conceptual truth that we are actually in an ABC and XYZ world. It would be compatible with MFP2's being a conceptual truth that the actual world is one of the bizarre humiliation-friendly worlds where it is not wrong to humiliate others just for fun. I will say more about this in what follows. For present purposes, the important point is that Cuneo and Shafer-Landau seem clearly to want a reading on which MFP2 is a conceptual truth regarding the actual world so described. They think it is a conceptual truth that we actually ought not to humiliate people just for fun. The indexical reading gives them this.

The trouble is that, on the indexical reading, the claim that the fixed point propositions are conceptual truths has problematic implications regarding what can be known *a priori*. Consider MFP2, for example. On Cuneo and

Shafer-Landau's account of conceptual truth, if it is a conceptual truth that [it is wrong to humiliate people just for pleasure in worlds like ours for beings like us], then it belongs to the essence of one or more of the concepts that constitute the proposition that, necessarily, anything that satisfies the complex concept of [humiliating a person just for pleasure in a world like ours for beings like us] also satisfies the concept of being wrong (2014: 410, with n. 27). Suppose for the sake of argument that it is wrong to humiliate people just for pleasure in all and only ABC and XYZ worlds. The problem is that it is not a conceptual truth that our world is an ABC and XYZ world. But on the indexical interpretation, this must be a conceptual truth, if MFP2 is a conceptual truth. For on Cuneo and Shafer-Landau's account of conceptual truth, if MFP2 is a conceptual truth, then the essence of the concept wrongness, or perhaps the essence of the complex concept of [humiliating a person just for pleasure in a world like ours for beings like us], or perhaps both, must be such as to necessitate that anything satisfying the complex concept also satisfies the concept of being wrong. But then this complex concept or the concept of wrongness or both must ensure that our world is an ABC and XYZ world. But this is not something that is settled *a priori* on the basis of relations among concepts. As I will argue next, it is a conceptual possibility that the actual world is one of the humiliation-friendly worlds where it is not wrong to humiliate others just for fun. Conceptual competence would not suffice to rule this out.

There are possible worlds in which a small dose of humiliation, when it is done just for pleasure, tends to help a person develop a virtuous kind of humility and tends to undermine any tendency to a vicious kind of arrogance. A small dose of humiliation might also contribute to a kind of psychological strength that would redound to a person's benefit. Call worlds like this, "Humiliation Friendly." It is arguable that it would be morally permissible to humiliate others in minor ways in Humiliation-Friendly worlds even if one does this for fun. So to make it plausible that it is wrong to humiliate others just for fun in worlds like ours, one needs the further thesis that our world is not Humiliation Friendly. But the proposition that our world is not Humiliation Friendly obviously is not a conceptual truth. This means that MFP2 is not a conceptual truth on the indexical interpretation. That is, it is not a conceptual truth that it is wrong to humiliate others just for fun in worlds like ours and for beings like us. For all that our concepts encode, our world might be Humiliation Friendly.

I cannot argue the point here, but I believe that analogous problems arise with respect to the other substantive moral claims on Cuneo and Shafer-Landau's list of fixed point propositions. If so, this is a fatal problem for the moral fixed points thesis, it seems to me, at least on the indexical interpretation.

Cuneo and Shafer-Landau anticipate an objection of this kind. Their response is to acknowledge that the essence of the concept wrongness does not directly guarantee the truth of empirical claims about our world or our

psychology, such as that our world is not Humiliation Friendly. Hence, the essence of the concept wrongness does not directly guarantee that it is wrong to humiliate others just for fun in worlds like ours for beings like us. They claim, however, that it might nevertheless be a "mediate" conceptual truth that it is wrong to humiliate others just for fun in worlds like ours for beings like us (2014: 433–435).

To illustrate the idea, they show that we can arrive, by a chain of reasoning, at the conclusion that the concept wrongness is such that it is a conceptual truth that only beings who are basically like us in worlds like ours are capable of performing actions that could be wrong. Say, for short, that it is a conceptual truth that if an action is wrong, it was performed by an Agent in an Agential world—where as a matter of fact, we are Agents and the actual world is an Agential world. And so, they seem to conclude, it is a conceptual truth that if an action is wrong, it was performed by a being like us in a world like ours. If this is correct, it follows that [if it is wrong to humiliate others just for fun, then this is wrong for Agents in Agential worlds]. This would be a conceptual truth. The problem, however, with Cuneo and Shafer-Landau's reasoning, is that it is not a conceptual truth that [we are Agents and our world is an Agential world]. Hence, it does not follow that it is a conceptual truth that [if it is wrong to humiliate others just for fun, then this is wrong for agents like us in worlds like this].

There is a further problem. For even assuming that [if it is wrong to humiliate others just for fun, then this is wrong for Agents in Agential worlds], it does not follow that [it is wrong for Agents in Agential worlds to humiliate others just for fun]. Humiliation-Friendly worlds are counter-examples.

In short, Cuneo and Shafer-Landau are right that there can be conceptual truths that concern our nature as Agents in an Agential world. For arguably it is a conceptual truth that wrong actions are done by Agents in Agential worlds. By contrast, the proposition that our world is not Humiliation Friendly is not a conceptual truth. It is not a conceptual truth that it is wrong to humiliate others just for fun in worlds like ours.

Perhaps, however, Cuneo and Shafer-Landau intend their fixed points thesis to be given the descriptions interpretation. To fill out the content of MFP2 on this interpretation, however, they would need to specify which facts about worlds and agents make a world Humiliation Friendly and then formulate a proposition about the circumstances in which humiliating others for fun is wrong that they could then argue to be a conceptual truth. Let us stipulate that any world in which humiliating others for fun would *not* be wrong is "Humiliation Friendly." Then their claim would need to be that it is a conceptual truth that [it is wrong to humiliate others just for fun *except* in worlds that are Humiliation Friendly]. This arguably is a conceptual truth, given our stipulation, but it is not substantive. It is trivial, analogous to the truth that wrongful killing is wrong. It would amount to the claim that it is wrong to humiliate others just for fun except in worlds where it isn't wrong. Could one get past this problem by offering a theory

of the humiliation friendly? Could we defend a theory, for some properties ABC and XYZ, according to which worlds with these properties are all and only the worlds in which it is wrong to humiliate people just for pleasure? The problem is that, at least as far as I can see, the concept of wrongness does not encode such a theory. It does not encode all the circumstances that would make a world be not Humiliation Friendly.

To conclude, I have argued that the fixed point propositions are substantive moral claims but that they are not conceptual truths. For it is not a conceptual truth that we are Agents and that our world is an Agential world. Further, it is not a conceptual truth that, for example, our world is not Humiliation Friendly. For these reasons, I think Cuneo and Shafer-Landau cannot achieve the ambitious goal of showing that the fixed point propositions are conceptual truths on the indexical interpretation. I cannot rule out in the same way the thesis that the fixed point propositions are conceptual truths on the descriptions interpretation. But this thesis may not be problematic for naturalism. It is not problematic for naturalism that it is a conceptual truth that wrongful killing is wrong. Similarly, it would not be problematic if, for example, it is a conceptual truth that [it is wrong to humiliate others just for fun except in worlds that are Humiliation Friendly].

7.   Conclusion

Avant-garde nonnaturalism attempts to avoid the familiar challenges to orthodox nonnaturalism by eschewing robust metaphysical or epistemological commitments. One such position is Conceptual Nonnaturalism. On this position, there is no need for the nonnaturalist to postulate nonnatural properties; nonnaturalism requires only that there are irreducible nonnatural moral *concepts* and substantive moral conceptual truths involving these concepts. In this way, it aims to avoid both the metaphysical and the epistemological challenges to orthodox nonnaturalism. In this essay, I have discussed the bold version of Conceptual Nonnaturalism that has recently been proposed by Cuneo and Shafer-Landau (2014).

I have reached a pessimistic conclusion about Cuneo and Shafer-Landau's account. Their theory of concepts and conceptual truth faces epistemological and metaphysical challenges that are akin to those faced by orthodox nonnaturalism. So it appears that their theory does not offer a significant improvement on orthodox nonnaturalism. The heart of their theory is the fixed points thesis, the thesis that the so-called moral fixed point propositions are substantive conceptual truths. I agree that these propositions are substantive—a philosopher who denied that there are moral properties would be committed to denying them—but it is implausible that they are conceptual truths. If they were conceptual truths, then the error theory would be conceptually false, but it is implausible that this theory is conceptually false. It is not plausibly a conceptual truth that there is a property of wrongness. Furthermore, once the moral fixed point propositions are

qualified to make them immune to bizarre counter-examples, it becomes plain that they are not conceptual truths. For it is implausible that these qualifications are built into our concepts. Indeed, on Cuneo and Shafer-Landau's view, we cannot even be sure that there is a single wrongness concept. There could be different moral conceptual truths relative to different concepts of wrongness.

The upshot is pessimistic and skeptical about the prospects for Conceptual Nonnaturalism. I do not claim to have shown that Conceptual Nonnaturalism is non-viable. But a viable version of Conceptual Nonnaturalism would need to rest on a theory of concepts and conceptual truths capable somehow of underwriting the thesis that there are substantive moral conceptual truths. If I am correct to think it implausible that it is a conceptual truth that there are moral properties, then it is equally implausible that there are substantive moral conceptual truths.[25]

Notes

1 I put so-called "quietist" theories in this camp. Among those who advocate forms of non-ontological nonnaturalism, as I understand the view, are Dworkin (1996), Skorupski (2010), and Scanlon (2014). I have discussed Parfit's views in Copp (2012; 2017). I have discussed Non-Realist Cognitivism in Copp (Forthcoming).
2 In earlier work, I offered a more detailed and cautious development of an epistemological characterization (Copp 2007b). Shafer-Landau opts in the end for an epistemological characterization (2003: 61). Parfit in effect proposes a similar account (2011: II, 306–307).
3 A neo-Russellian about propositions might object that if there is not a property of F-ness, then there are no propositions about the instantiation of F-ness, so there are no false propositions of this kind. On such a view, for example, if there is not a property of wrongness, then there are no propositions about the wrongness of things, so there are no false propositions of this kind. Cuneo and Shafer-Landau (2014: 411–412) hold, however, that, except in special cases, propositions are constituted by concepts, so for present purposes we can set aside the neo-Russellian objection. There is presumably a *concept* of wrongness.
4 One might worry that I am assuming some form of cognitivism about moral judgments according to which the logical form of the judgment that lying is wrong is the same as the logical form of the judgment that lying is widespread. This is not so. I will say more about this issue in a note in Section 5. Let me here simply point out that my argument is compatible with a "minimalist" account of moral properties, which is compatible in turn with a "quasi-realist" and "quasi-cognitivist" account of moral judgment (Blackburn 2006). According to a minimalist view, the claim that there is a property of wrongness is equivalent to the claim that some actions are (or would be) wrong. On this view, of course, if there is no property of wrongness, it is not the case that lying is wrong.
5 A skeptic might not want to *assert* that wrongful killings are wrong, however, since, at least in certain contexts, doing so might seem to presuppose that there is such a thing as wrongness. Yet she can agree that, if there are any wrongful killings, they are wrong, and, it seems to me, she need not *deny* that wrongful killings are wrong.
6 For this use of the "over-and-above" locution, see Enoch (2011: 101–102).
7 Cuneo and Shafer-Landau's "Minimal nonnaturalism" is a version of Conceptual Nonnaturalism that does deny this (see Cuneo & Shafer-Landau 2014: 405).

8 As I will eventually explain, it is not obvious that a naturalist would need to deny the Core Claim. For now, I ignore this worry.

9 Darwall (2006) advocates a similar view.

10 Here I am using "fact" to refer to a state of affairs, not merely to a true proposition. Cuneo and Shafer-Landau (2014: 414) refer to this as the "familiar correspondence" view about the nature of facts.

11 In a similar way, someone who holds that water is H_2O can nevertheless agree that the proposition that the glass is full of water is distinct from the proposition that the glass is full of H_2O molecules. She can deny that the concept of water is analyzable in terms of the concept of H_2O molecules, for her claim is about the property of being water, not about the concept.

12 Cuneo and Shafer-Landau (2014: 418–422) contend that there is an argument, the "Reversal Argument," that takes us from the Embellished Core Claim to a thesis that is a close relative of Robust nonnaturalism. If they are correct, the moral fixed points thesis and the amended Core Claim are entry points to a view that naturalists certainly would want to deny.

13 This leaves it open whether there are properties that are not represented by concepts.

14 According to their Minimal nonnaturalism, these would be natural properties.

15 Cuneo and Shafer-Landau might think there is a panoply of concepts, for on their view, the existence of a concept does not depend on our having discovered it or tokened it (2014: 410). It is not clear whether they think that every property is represented by a concept.

16 Strictly speaking, since propositions are constituted at least in part by relevant concepts, if there are different wrongness concepts, then there are correspondingly different propositions to the effect that humiliating people just for pleasure is "wrong." If there are different wrongness concepts, there are different fixed point propositions. MFP2 comes in as many "flavors" as there are wrongness concepts. I ignore this complication in the text.

17 Recall that, in Cuneo and Shafer-Landau's view, the moral fixed point propositions are constituted by concepts (2014: 403) and "are true in virtue of the essences of their constituent concepts" (2014: 411–412).

18 Richard Joyce (2001) has also defended the view that there are no moral properties.

19 One might worry that perhaps the claim that lying is wrong merely *presupposes* that there is a property of wrongness, and does not *entail* that there is. If so, one might think, the error theorist is not committed to thinking that the claim that lying is wrong is false, but only to thinking that it is mistaken or not true. I will ignore these subtleties here. For my purposes, it is enough if an error theorist is committed to denying a claim in any of these ways—by holding it to be false, or mistaken, or not true. If the error theory entails that a conceptual *truth* is false, or mistaken, or not true, then the theory is false.

20 I discussed this issue before in Copp (2007d: 126–127).

21 As we just saw, they say that "the concept 'being wrong' is such that, if anything satisfies a concept such as 'recreational slaughter,' then it also satisfies the concept 'being wrong'" (2014: 414, n. 33). In their view, the conceptual truth is anchored in both concepts. MFP1 is not relevantly analogous to the proposition that wrongful killing is wrong.

22 To be sure, an expressivist can perhaps coherently claim that recreational slaughter is wrong while denying that there is a property of wrongness in anything but a minimalist sense (Blackburn 2006). But Cuneo and Shafer-Landau are not proposing an expressivist or non-cognitivist analysis of moral judgment, so we do not need to investigate this gambit.

23 According to Cuneo and Shafer-Landau's Minimal nonnaturalism, this property would be a natural property.

24 There is a complication that I want to set aside for reasons of simplicity. To illustrate, consider the sentence, "I am here now." Arguably, to determine what proposition is expressed by this sentence, we need to specify a context such that, given the specified context, the sentence expresses a proposition about the location of the contextually relevant person at the contextually relevant time. Put abstractly, the complication is that, arguably, indexicals are terms that figure in *sentences*, where the *propositions* expressed by those sentences at relevant contexts are not in any interesting sense indexical. Cuneo and Shafer-Landau have a different view, however. For them, propositions are constituted by concepts, and they hold that there are indexical concepts (2014: 410, n. 25). They seem to hold that sentences containing indexical terms express propositions that are partly constituted by corresponding indexical concepts. The indexical interpretation discussed in the text assumes that this is their view.

25 I am grateful to Terry Horgan, Diego Machuca, Laura Schroeter, François Schroeter, and Mark Timmons for helpful discussion of the issues raised here.

References

Benacerraf, P. 1965. "What Numbers Could Not Be," *The Philosophical Review* 74: 47–73.

Blackburn, S. 2006. "Anti-Realist Expressivism and Quasi-Realism." In D. Copp (ed.), *The Oxford Handbook of Ethical Theory*, 146–162. New York: Oxford University Press.

Copp, D. 2007a. *Morality in a Natural World*. Cambridge: Cambridge University Press.

———. 2007b. "Why Naturalism?" In Copp 2007a, 33–54.

———. 2007c. "Moral Naturalism and Self-Evident Moral Truths." In Copp 2007a, 93–112.

———. 2007d. "Moral Necessities in a Contingent World." In Copp 2007a, 113–150.

———. 2012. "Normativity and Reasons: Five Arguments from Parfit against Normative Naturalism." In S. Nuccetelli & G. Seay (eds.), *Ethical Naturalism: Current Debates*, 24–57. Cambridge: Cambridge University Press.

———. 2017. "Normative Naturalism and Normative Nihilism: Parfit's Dilemma for Naturalism." In S. Kirchin (ed.), *Reading Parfit on What Matters*, 28–53. London: Routledge.

———. Forthcoming. "A Semantic Challenge to Non-Realist Cognitivism."

Cuneo, T., & Shafer-Landau, R. 2014. "The Moral Fixed Points: New Directions for Moral Nonnaturalism," *Philosophical Studies* 171: 399–443.

Darwall, S. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.

Dworkin, R. 1996. "Objectivity and Truth: You'd Better Believe It," *Philosophy and Public Affairs* 25: 87–139.

Enoch, D. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.

Gibbard, A. 2003. *Thinking How to Live*. Cambridge: Harvard University Press.

Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.

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

Parfit, D. 2011. *On What Matters*, 2 volumes. Oxford: Oxford University Press.

———. 2017. *On What Matters*, Volume 3. Oxford: Oxford University Press.

Railton, P. 2003. *Facts, Values and Norms: Essays toward a Morality of Consequence*. Cambridge: Cambridge University Press.

Scanlon, T. M. 2014. *Being Realistic about Reasons*. Oxford: Oxford University Press.

Shafer-Landau, R. 2003. *Moral Realism: A Defence*. Oxford: Clarendon Press.

Skorupski, J. 2010. *The Domain of Reasons*. Oxford: Oxford University Press.

6 The Phenomenology of Moral Authority

Terry Horgan and Mark Timmons

1. Introduction

Error theory in meta-ethics is motivated by skepticism about the claim that there are properties and relations instantiated in the world of the kind supposedly posited in moral experience, moral thought, and moral discourse. Perhaps the most serious such concern focuses on the aspect of *authority*, which seems central to moral experience.[1] Certain actual or potential actions (including refrainings) are experienced as being *morally fitting* in light of pertinent non-normative factual considerations—considerations which thereby are experienced as *reasons* for so acting. This fittingness-connection between non-normative considerations that constitute reasons and actions normally is experienced as having a governing authority over oneself that is independent of pre-existing desires and conventional social norms. In that respect, this authority is experienced as categorical—and thus as "inescapable" in its rational-justificatory import (even when one flouts it by acting contrary to what one experiences to be fitting). Another way to put this point is to say that the non-normative considerations themselves are experienced as being categorically authoritative, vis-à-vis the action that they render fitting.[2]

When one finds oneself making a moral judgment—e.g., that a certain kind of action is morally obligatory in the currently prevailing circumstances—one's experience is as-of that action's having this moral status in virtue of its fittingness in the pertinent circumstances. As we will put it, moral judgments are experienced as being *fittingness-based*. The aspect of categorical authoritativeness thereby accrues to the moral judgment itself, since the judgment is experienced as being based on a categorically authoritative fittingness-connection.

Those who embrace error theory because of the categorical-authoritativeness aspect of moral experience do so on the basis of the contention that, if putative relations or properties that are inherently categorically authoritative were really instantiated in the world, then this would be intolerably metaphysically queer from the perspective of any broadly naturalistic metaphysical outlook. This was one dimension of J. L. Mackie's famous "queerness

argument" for error theory (Mackie 1977). Jonas Olson (2014) has recently distinguished several distinct strands in Mackie's argument that Olson claims were not sufficiently distinguished by Mackie himself. Olson argues on one hand that several of those strands are not terribly persuasive, but on the other hand that the argument from the queerness of categorical authoritativeness is sound. A similar claim is made by Garner (1990) and Joyce (2001) who, like Olson, embrace error theory on the basis of this specific argument.

The categorical-authoritativeness aspect of moral experience sometimes has been invoked in support of non-naturalist moral realism. The claim is that this meta-ethical position is the only one that can adequately accommodate this experiential feature—as opposed to naturalist moral realism, or expressivism, or (of course) error theory. To "accommodate" such a feature—as we employ this word here and throughout—is to do two things: first, to acknowledge it as an inherent feature of moral experience, thought, and discourse; and, second, to treat it as a feature that does not ubiquitously misrepresent the world—rather than being ubiquitously non-veridical, as error theorists like Olson, Garner, and Joyce contend. This line of argument was developed at some length by the late Jean Hampton in an incomplete, posthumously published book (Hampton 1998); similar argumentation is deployed by Nagel (1986) and Enoch (2011).

Both sides in this dispute embrace the following two theses:

(1) *Categorically Authoritative Fittingness*: ordinary moral experience has an inherent aspect of categorically authoritative fittingness.
(2) *Ontological Purport*: the aspect of categorically authoritative fittingness in ordinary moral experience purports to represent both (*i*) instantiated fittingness-relations that are inherently categorically authoritative, and (*ii*) fittingness-based instantiated moral properties that inherit this categorical authoritativeness.

Hampton affirms the existence of instantiated, categorically authoritative, moral fittingness-relations and moral properties. Mackie and companion error theorists instead embrace the following skeptical thesis:

(3) *Ontological Moral-authority Skepticism*: no properties or relations are ever instantiated that are inherently categorically authoritative.

Mackie and companions thereby embrace the conjunction of (1), (2), and (3)—a view we will call *standard moral-authority skepticism*.

Our goal here is to articulate and motivate a version of meta-ethical expressivism, which we call 'cognitivist expressivism' and which seeks to acknowledge and accommodate thesis (1). This position embraces both (1) and (3)—a view we will call *moderate moral-authority skepticism*. However, in opposition both to Hampton-style non-naturalist moral realism

and to error theory, it also affirms the negation of (2) and instead embraces an expressivist, metaphysically irrealist, construal of the aspect of categorically authoritative fittingness in moral experience.

Standard moral-authority skepticism is skeptical about ordinary moral experience—and by extension, ordinary moral thought and discourse. It treats these as embodying an ontological error. By contrast, the moderate moral-authority skepticism embraced by cognitivist expressivism does not impugn ordinary moral experience at all (or ordinary moral thought and discourse). Rather, it ascribes error only to certain meta-ethical views, viz., those that embrace the conjunction of theses (1) and (2). Moreover, since moral-authority expressivism embraces thesis (1), the specific meta-ethical claim that it treats as erroneous is thesis (2). (Although thesis (2) presupposes thesis (1), the contention of cognitivist expressivism is that (1) does not entail (2)—and, moreover, that (1) is true whereas (2) is false.)

Perhaps the most powerful way to argue in favor of the conjunction of theses (1) and (2) is by appeal to phenomenology, i.e., to considerations about "what it is like" to undergo moral experience. Jean Hampton articulated such an argument in some detail in Chapter 3 of Hampton (1998), in a way that we think could be happily embraced not only by non-naturalist moral realists like herself but also by moral error theorists like Mackie, Olson, and others. We therefore focus, in Section 2, on her discussion. In Section 3 and based on our discussion of Hampton, we consider a phenomenological argument for non-naturalism that is intended to favor this view over error theory, but we also explain how an error theorist should respond. It might then seem that non-naturalism and error theory are the only views that recognize the pertinent phenomenology of moral authority. However, in Section 4, we sketch the elements of our favored meta-ethical view we call 'cognitivist expressivism' and proceed to explain how it not only recognizes the phenomenology of moral authority just as thoroughly as do non-naturalism and error theory, but that cognitivist expressivism, unlike error theory, can *accommodate* this phenomenology (as does non-naturalism). Central to our defense of this claim is what we call the 'neutrality thesis,' according to which one cannot determine on the basis of introspectionist methodology, employed in phenomenological arguments, whether concrete moral experience purports to represent the instantiation of the sorts of fittingness-relations mentioned in *ontological purport*. What the neutrality thesis makes clear is that an expressivist view like ours can happily embrace *categorically authoritative fittingness*, yet reject *ontological purport*. And so we argue that our view can accommodate the phenomenology of moral authority. We conclude, in Section 5, with a brief cost-benefit comparison of non-naturalism and error theory with our cognitivist expressivism, calling attention to the comparative advantages of this latter view.

2.   Hampton on Moral Authority

In her third chapter, entitled "Reason's Authority," Hampton's stated goal is to argue that there is something about ethics "that appears to make it scientifically problematic," viz., "a certain thesis held by moral objectivists about moral norms and the reasons they generate that fails to pass scientific muster" (1998: 83). Her argument is phenomenological, focusing on introspectively salient aspects of moral experience. The argument proceeds in two stages. In the first stage, she offers what she says is a "minimalist and metaphysically neutral" initial characterization of how the normativity of moral reasons is experienced (1998: 83). In the second stage, she describes two metaphysically different ways of "embellishing" the initial characterization (1998: 83), and she argues that only the second way—a version of non-naturalist moral realism—comports with the phenomenology of moral-reasons experience. In this section, we briefly summarize her argument, in a way that closely follows her own text.

2.1.   *The First Stage*

Hampton begins by saying that the notion of authority in question is not decisiveness, because the reasons in support of a particular act can carry authority even if they are outweighed by reasons in support of some other act; and that the pertinent notion is not strength of motivational commitment either, because the reasons in support of a particular act can be experienced as stronger, qua reasons, than considerations upon which one chooses to act. Rather, she says,

> a theory of the authority of norms tries to explain *what it means* for a norm to be "applicable" to us. The word 'applicable' is a poor one, because norms don't merely "apply" to us, they *direct* us. Indeed, we use all sorts of words to elaborate on this applicability: besides "authority," we speak of a norm's "prescriptivity" or its "obligatory force" over us, its "compelling nature" or its "pull," its status as in "order" or a "command" (and not a mere "suggestion"). . . . That is, normative authority . . . presupposes that it is correct to say that it specifies a reason for x-ing for an agent.
>
> (1998: 88)

Here she does not say explicitly what she takes a "norm" to be, and for present purposes this does not much matter. The crucial thing presently, we take it, is this: to *experience* a norm "as specifying a reason for x-ing for an agent," where consideration C is the specified reason, is to experientially *regard* consideration C as a reason for x-ing for an agent. (Presumably, the agent in question might or might not be oneself. Often enough, one experiences a moral reason as universalizable: a reason for x-ing for *anyone* who might be in the pertinent circumstantial situation.)

She points out that

> this compelling quality of reasons is not . . . the same as the feeling of liking or approving the directive of a reason. It is easy to give examples of people who know they have a reason to x, and who not only do not like the action that the reason directs, but even despise it.
>
> (1998: 89)

She also points out that not all reasons have the same kind of directive force: "In particular, some reasons command us, and thereby give us mandates, and others direct us in ways that indicate permissions, rather than commands" (1998: 90). Nonetheless, "what permissive and mandatory reasons have in common, such that they are all *reasons*, is their authority—the sense in which they have for us a 'compelling rightness'" (1998: 91).

She takes all this to be phenomenological description, a matter of isolating "the distinctive 'feel' of reasons" (1998: 93). She also takes it to be so far metaphysically neutral; she denies that she has meant "to suggest that this compelling quality must be understood as somehow 'in the world,' and thus a part of our reality" (1998: 93). But next comes stage 2 of her argument, in which she articulates and compares two competing theses about normative authority—specifically, about moral normative authority—each of which she considers metaphysical.

2.2. *The Second Stage*

The first thesis she considers might naturally be regarded as compatible with a broadly naturalistic metaphysical worldview, and as not being "scientifically problematic." As a prelude to introducing it, she begins with the following observations:

> Consider a norm in the sport of dressage, requiring owners of horses performing dressage tests in a competition to present the horse in the show ring with its mane braded. It purports to give those who recognize it a reason to act in a certain way. . . . [T]o explain the authority of this or any similar norm, such as a rule in baseball, or a norm of etiquette, or a norm about taste in foods, we would tell a story locating the social forces that generated the norm, and the psychological responses to those forces by certain people that give these norms their power. In this view, the norm is the (mere) invention of particular human beings by virtue of their interests, and the sense such people have that a norm authoritatively applies to them . . . is entirely a matter of social and psychological contingencies. . . . This sort of norm, which I will call culture-dependent, is ubiquitous, and includes norms of etiquette, rules of various sports, and ideals of physical beauty.
>
> (1998: 94)

She now introduces the first thesis as follows, by way of contrast with norms that are obviously and unproblematically culture-dependent: "some theorists believe that *all* norms, including all moral norms, are culture-dependent. Those who believe this explain the authority of all norms as (what I will call) a 'psycho-social' phenomenon" (1998: 94). (She is here using the expression "culture-dependent' broadly enough to encompass psychological factors that might be innate or otherwise deeply ingrained in human nature, as well as factors resulting from socialization.) She goes on to characterize a theoretical burden that she says any psycho-social approach must shoulder:

> Of course, a sophisticated version of this account is going to want to explain why "common sense" tells us there is a difference between (what clearly seem to be) "artificial" norms such as monopoly rules, dressage standards, and codes of etiquette on the one hand, and moral norms such as "do not murder" and "treat all people as equals." There are a number of ways to construct such a sophisticated view, and each of these I take to be a variant of the psycho-social thesis about normative authority.
>
> (1998: 94–95)

One variant is meta-ethical expressivism, about which she says the following (with specific reference to Allan Gibbard):

> In this view, the apparently "special" authority of moral norms is (in some way) a function of how people accept them. That is, our sense of their necessity is a product of our commitment to them (but not the other way around). . . . In this view, our sense that reasons emanating from norms have a certain compelling rightness that takes either a mandatory or a permissive form is a (mere) feeling that is attached to directives, such that we accept them, feel compelled by them, and are motivated by them.
>
> (1998: 95)

Another variant is error theory, about which she says the following (with specific reference to J. L. Mackie):

> This view . . . explains the authority of reasons as deriving from people's acceptance of a certain kind of theory. . . . The theory that has been current in our society is the objectivist's account of authority, as something that is independent of human psychology and culture. . . . But in fact this theory is in error, because there is no such objective authority. . . . Hence, in this view, people believe that the authority of many norms is objective because they accept a false theory—a theory whose acceptance may nonetheless be highly attractive to them (for psychological reasons).
>
> (1998: 95–96)

Concerning naturalist moral realism, she says the following (with specific reference to Richard Boyd): "The most that Boyd's position allows is that we can say that something is a reason for an agent if her psychology is such that she happened to regard it as such" (1998: 102). This comment appears to bring naturalist moral realism too under her rubric of "psycho-social" meta-ethical positions. For, on this view, presumably, the difference between the phenomenology of reasons that are clearly culture-dependent on one hand, and the phenomenology of moral reasons on the other hand, is psycho-socially explainable: whereas people typically experience being free to choose whether or not to engage in social practices governed by norms that are clearly culture-dependent, nevertheless most people—because of contingent psychological factors, either innate or enculturated or both—do not experience such optionality with respect to moral norms and the reasons they engender.

She summarizes as follows what she takes to be the common thread that runs through the various respective versions of the psycho-social metaphysical approach:

> All these versions of the psycho-social thesis accept the same basic strategy for explaining the authority of reasons—that authority is understood to be merely in the head (explicated as a feeling, or a cognitive state, or a theoretical belief), and its origins are explicable by virtue of human psychology, human biology, and/or human sociology.
>
> (1998: 96)

She turns next to a second metaphysical "elaboration" of the features of moral-authority experience that she described in the first stage of her argument—an alternative elaboration that she takes to be incompatible with the psycho-social thesis. She calls it the "objectivist" thesis, which she characterizes this way:

> Those who are normative objectivists maintain that some norms (but not all norms—for example, not norms of dressage) are examples of what I will call culture-independent or objective norms. The authority of these norms is supposed to be independent of social and psychological contingencies. . . . [O]bjectivists argue that such (independent) authority is the reason society has (or ought to have) such norms as part of its culture, and they insist that no matter the facts of our society or our psychology, we ought to recognize, accept, and obey them.
>
> (1998: 96)

The notion of objective authority, she says, figures in the objectivist view in a number of ways:

> First, and most importantly, . . . such authority is "outside" the agent, and that to which she is responding when she says that she understands

that she ought to act from them. . . . [T]he authority is not the invention of the agent, nor of human communities, but something to which agents and human communities respond.

> Second, the moral objectivist assumes that the notion of authority is one that human beings can "see" or (in some way) discover. . . . [W]e usually "feel" or "comprehend" its (objective) authority, which means experiencing a sense of its pull, such that we take it to be something that we are in certain circumstances bound to act upon.

> Third, the moral objectivist claims that having felt this authority, it is—at least sometimes—an authority for the sake of which we can act, so that it is motivationally efficacious.

<div align="right">(1998: 98–99)</div>

She further elaborates the aspect of "outsideness" as involving a distinctive kind of necessity, about which she says:

> Let us say that moral reasons generated by objectively authoritative moral norms are necessary in the sense that their *governance* over us is inescapable. And by 'inescapable' here I mean that these reasons "apply" to us "no matter what." According to this way of thinking about objective authority, no matter what we may do or think, we are directed by these reasons—ether in the form of permissions or in the form of mandates. And the governance is inescapable or necessary because there is no way that we can throw it off, or change it by our actions, beliefs, or social systems. . . . So understood, normative necessity is still a metaphysical concept because it is supposed to hold regardless of whether or not we know about it or are aware of it. It is just not the metaphysical concept that is usually referred to by the term "necessity."

<div align="right">(1998: 105–106)</div>

And she urges that the "outsideness" of moral normative authority, with its aspect of necessity or inescapability, is central to the actual phenomenology of moral-reasons experience:

> This way of thinking about authority is, I think, closest to the way that the authority of reasons *feels* to us—that is, it approximates what the experience of "having a reason" is like for those who understand and act from reasons. . . . [R]easons feel like orders—strong in the case of mandates, weak in the case of permissions, but directives nonetheless, with an inescapable rightness about them.

<div align="right">(1998: 106)</div>

Although she does not say so explicitly, her overall discussion of moral authority clearly implicates—especially in light of the just-quoted passage—that the objectivist thesis fits people's actual moral-authority phenomenology better than does the psycho-social thesis.

Let us now itemize, for purposes of subsequent citation, various interconnected features that Hampton has identified as elements of the phenomenology of categorically authoritative moral fittingness. For simplicity, we restrict attention to reasons that one experiences as morally *requiring* a certain action, and as requiring such an action *by oneself*. (Her discussion can be generalized to cover reasons that one experiences as favoring a certain action without requiring it, and to cover reasons that one experiences as pertaining to other persons—or to anyone who might find oneself in certain circumstances.) And we focus, as does Hampton, upon experiences of *non-normative* considerations as being reasons for a certain kind of action—non-normative considerations in virtue of which such an action is experienced as being categorically-authoritatively fitting. The key phenomenological elements of such fittingness-experience are these:

- *Pull*: such reasons are experienced as pulling one toward a certain specific action (perhaps a refraining).
- *Independence*: the pull of such reasons is experienced as being independent of pre-existing desires or human conventions.
- *Source*: the source of the independence is thus experienced as "outside" oneself.
- *Inescapability*: such reasons are experienced as inescapable.
- *Grip*: such reasons, when not experienced as being outweighed by other reasons of the same kind that pull toward some incompatible action, normally exert an involuntary, categorically authoritative, experiential grip upon oneself—a grip toward performing the action toward which one is pulled.
- *Motivation*: such experiences typically motivate one to act accordingly (although their motivational strength can be outweighed by other psychological states such as pre-existing desires).

3. Incorporating the Phenomenology of Moral Authority into Meta-Ethics

The categorical-authoritativeness aspect of moral phenomenology, which includes each of the elements just summarized, is important to acknowledge as a key feature of moral experience, moral judgment, and moral discourse; it needs to be taken into account by any viable meta-ethical position. *Prima facie*, this phenomenology favors the position that Hampton advocates in her book—viz., a version of non-naturalist moral realism that posits instantiated, categorically authoritative, in-the-world properties and relations. (Here and henceforth, we use the expression 'in-the-world' to signal an ontologically committal use of talk about properties, relations, states of affairs, etc.—as opposed to a minimalist, ontologically "lightweight" way of using such talk that could be deployed even by those who are irrealists about moral metaphysics, such as error theorists and expressivists.)

Hampton's phenomenological description of moral-reasons experience strikes us as plausible and compelling. We will not contest it here. Rather, we will focus on the following question: given that the categorical-authority aspect of moral experience has the phenomenological features she describes, can these features be plausibly incorporated into any meta-ethical position(s) other than non-naturalist moral realism?

In the present section, we address this question dialectically. We begin by formulating an argument—we think implicit in Hampton's discussion of moral authority—in support of a negative answer. We then set out a plausible-looking critique of this argument that can be offered by an error theorist, together with a modified version of the argument that the error theorist can embrace. This modified version generates the conclusion that only error theory and non-naturalist moral realism are compatible with the phenomenology of categorical moral authority—a conclusion that feeds into the variant of the metaphysical "queerness" argument for error theory that is embraced by Olson and fellow error theorists. We then sketch a potential dialectical response to all this on behalf of expressivism. The expressivist can embrace the error theorist's critique of Hampton's own implicit phenomenological argument, and can go on to repudiate both Hampton's version of the argument and the error theorist's modified version of it—by claiming that both arguments rest on a false presupposition, viz., proposition (2) in the introduction to the present chapter, the contention we labeled "ontological purport." This will leave the expressivist with the task of accommodating the categorical-authoritytitativeness aspect of moral phenomenology in a way that eschews proposition (2); that will be the business of Section 4.

3.1. A Phenomenological Argument Favoring Non-Naturalist Moral Realism

Implicit in Hampton's discussion of moral authority is an argument for the conclusion that the only phenomenologically viable meta-ethical position is a version of non-naturalist moral realism that treats categorically authoritative moral fittingness as a genuine, in-the-world, ontological phenomenon. This argument, if sound, favors non-naturalist realism because incorporating the phenomenology of moral authority is an important desideratum for a meta-ethical theory. The argument can be formulated as follows. (Hereafter we use the label 'Alternative' to cover any meta-ethical position other than Hampton-style non-naturalist moral realism; and we use the label 'non-naturalist moral realism' only for Hampton-style versions of this position, i.e., versions that treat categorically authoritative moral fittingness as an in-the-world phenomenon. We also speak of the authority of *reasons* rather than the authority of fittingness or the (fittingness-based) authority of moral judgments, because this cleaves more closely to Hampton's own preferred terminology.)

Argument 1

1. On any Alternative meta-ethical account, the authority of moral reasons is culture-dependent and thus contingent, i.e., it is dependent upon certain contingent psychological and/or social factors.
2. But moral reasons are ordinarily experienced as being authoritative in a way that is culture-independent and thus necessary (what we earlier called the inherent aspect of categorically authoritative fittingness).

Therefore,

3. No Alternative meta-ethical account can accommodate the phenomenological categorical-authoritativeness aspect that is inherent to ordinary moral-reasons experience. (From 1–2)
4. A phenomenologically viable meta-ethical position must accommodate this aspect.
5. Non-naturalist moral realism accommodates it.

Therefore,

6. The only phenomenologically viable meta-ethical position is non-naturalist moral realism. (From 3–5)

Recall that "accommodating" the categorical-authoritativeness aspect of moral-reasons experience requires not only acknowledging it, but also treating it as a feature that does not ubiquitously misrepresent the world.

This argument has considerable *prima facie* plausibility. Nonetheless, an advocate of meta-ethical error theory will not be convinced—and should not be, for the following reasons.

3.2. An Error-Theoretic Response

Two distinct uses of expressions like 'the authority of moral reasons' should be clearly distinguished. On one hand is an *ontologically committed* use (as we will put it, deploying Quinean terminology), which presupposes that moral authority is an in-the-world, ontologically real, phenomenon. On the other hand is an *ontologically neutral* use (as we will put it), which, while acknowledging the moral-authoritativeness aspect of moral experience and thought and discourse, remains neutral about whether or not this experiential aspect successfully represents a genuine in-the-world phenomenon—as opposed to merely *purporting* to represent such a (putative) phenomenon.

Argument 1 can be read in two ways, corresponding to these two distinct uses of the pertinent terminology. On an *ontologically committed* reading, the argument presupposes that moral authority is a genuine in-the-world phenomenon. The crux of the argument, given this presupposition, is that

Alternative accounts of this phenomenon distort its real nature. These accounts treat its governing power over us as obtaining only contingently, whereas the phenomenology of moral authority reveals that it is actually necessary and inescapable. Such inescapability is a counterfactual matter: a moral reason would continue to govern us even if the pertinent, psycho-social, contingencies did not obtain.

On an *ontologically uncommitted* reading, on the other hand, the argument does not presuppose that the authoritativeness aspect of moral experience represents, successfully, a genuine in-the-world phenomenon. Rather, the expression 'moral authority' is to be understood as characterizing a *putative* in-the-world phenomenon which the authoritativeness aspect of moral experience *purports* to represent. The crux of the argument, on this reading, is that Alternative accounts misdescribe this phenomenological aspect itself. These accounts treat it as an aspect as-of being contingently normatively governed, whereas in fact it is an aspect as-of being necessarily, inescapably, normatively governed.

With all this in place, the advocate of moral error theory can now reply to the argument as follows. On the ontologically committed reading, the argument is question-begging against error theory, because it presupposes what the error theorist explicitly denies—viz., that there is such a genuine in-the-world phenomenon as categorical moral authority. On the onto-logically uncommitted reading, on the other hand, premise 4 is mistaken, because another viable way for a meta-ethical theory to come to terms with the phenomenological aspect of categorical authoritativeness is to *acknowledge it but treat it as non-veridical*—which is precisely what error theory does.

So on either reading, the error theorist can and should repudiate the argument. Also, whether or not one embraces error theory, the availability of this response seriously undercuts the argument's initial plausibility. Thus, advocates of various other Alternative views can accept the error theorist's critique of Argument 1—and should, we say. (This includes expressivists, of course.) But there is another phenomenological argument looming, similar in spirit to Argument 1, which error theorists can happily embrace. We turn next to that.

3.3.   A Phenomenological Argument Favoring Error Theory

The error theorist can—and should, we say—re-tool Argument 1 into a new argument whose conclusion is that there are two, and only two, phe-nomenologically viable meta-ethical views, viz., non-naturalist moral real-ism and error theory. This argument, if sound, indirectly favors error theory over all its competitors—because (a) no other Alternative theories turn out to be phenomenologically viable, and (b) it is very plausible that the puta-tive phenomenon of in-the-world categorical authoritativeness is intolerably metaphysically queer. The argument can be formulated as follows.

Argument 2

1. On any Alternative meta-ethical account, the authoritativeness aspect of moral-reasons experience is culture-dependent and thus contingent, i.e., it is dependent upon certain contingent psychological and/or social factors.
2. But moral reasons are ordinarily experienced as being authoritative in a way that is culture-independent and thus necessary (what we earlier called the inherent aspect of categorically authoritative fittingness).

Therefore,

3. No Alternative meta-ethical account can accommodate the phenomenological categorical-authoritativeness aspect that is inherent to ordinary moral-reasons experience. (From 1–2)
4. A phenomenologically viable meta-ethical position must either accommodate this aspect or treat it as erroneous.
5. Non-naturalist moral realism accommodates it.
6. Only error theory treats it as erroneous.

Therefore,

7. The only phenomenologically viable meta-ethical positions are non-naturalist moral realism and error theory. (From 3–6)

This argument has considerable *prima facie* plausibility, and is not at all impugned by the reasons given in Subsection 3.2 for questioning Argument 1. It poses a serious threat to all Alternative views besides error theory, expressivism included.

3.4. A Partial Expressivist Response

Expressivism can fend off this threat, we say. How? In this subsection, we offer a partial answer, which by itself will be incomplete because it abstractly outlines an expressivist approach to categorical-authoritativeness phenomenology whose concrete details still need filling in. Those details rwill be supplied in Section 4.

An expressivist, we suggest, should challenge the inference in Argument 2 from lines 1 and 2 to the preliminary conclusion 3. The challenge goes as follows. In order for the necessity/inescapability aspect of categorical-authoritativeness phenomenology to be incompatible with a contingent, "psycho-social" explanation of this phenomenological feature, this putative incompatibility would have to be based on the possession, by the phenomenology, of *ontological purport*. That is to say, it would have to be the case that statement (2) in this chapter's introduction—which we repeat here—is true:

(2) *Ontological Purport*: the aspect of categorically authoritative fittingness in ordinary moral experience purports to represent both (*i*) instantiated

fittingness-relations that are inherently categorically authoritative, and (*ii*) fittingness-based instantiated moral properties that inherit this categorical authoritativeness.

But an expressivist can, and should, deny (2)—and therefore should reject as unsound the initial subsidiary inference in Argument 2. The *occurrence* of moral-authoritativeness phenomenology in human experience can be a contingent, psycho-socially explainable, phenomenon even if that phenomenology is not experientially *as-of* normative governance that obtains only contingently. Moreover, and assuming the absence of in-the-world, categorically authoritative, moral-fittingness relations, moral phenomenology can have the aspect of *categorical* normative governance without being in error—provided that (2) is false, and thus that moral phenomenology *does not purport to represent* putative in-the-world, categorically authoritative, relations or properties.

So far, so good—but also so partial. The task facing the expressivist is to give a positive, expressivistic, account of moral experience that accommodates its inherent phenomenological aspect of categorically authoritative fittingness in a manner consistent with the denial of (2). We turn next to that, with specific attention to Hampton's characterization of moral-authority phenomenology.

4. An Expressivist Account of Categorical Authoritativeness

Arguments 1 and 2 are both phenomenological arguments: they rest on the contention that the aspect of categorically authoritative fittingness is an introspectibly discernible feature of moral experience. This contention we happily grant. Our aim here is not to dispute it, but rather to argue that it can be smoothly accommodated within an expressivist meta-ethical position.

We do, however, dispute claim (2). We contend that (2) is neither a claim that introspection itself directly and reliably reveals to be true, nor a claim that provides the only viable potential explanation of those elements of categorical-authoritativeness phenomenology that themselves are reliably discernible by direct introspection. We will argue that the version of expressivism we ourselves espouse can smoothly accommodate the directly introspectively discernible elements of categorical-authoritativeness phenomenology, thereby yielding a viable explanation of them that eschews both non-naturalist moral realism and moral error theory.

We will not be claiming that our proposed expressivist treatment of the phenomenological aspect of categorically authoritative fittingness can be reliably recognized as correct solely on the basis of direct introspection. Nor will we be claiming that our expressivist treatment qualifies, abductively, as the clearly best explanation of categorical-authoritativeness

phenomenology. On the contrary, we espouse a *Neutrality* thesis (as we call it), according to which direct introspection by itself cannot reliably determine whether claim (2) is true or false. We grant, therefore, that non-naturalist moral realism and error theory provide alternative, *prima facie* viable, potential explanations of the pertinent phenomenology. Insofar as one restricts oneself to phenomenological considerations—either the direct deliverances of introspection, or abductive considerations concerning the potential explanation of introspective phenomenological data—we contend only that the expressivist approach we will describe below constitutes *one prima facie* viable way to theoretically embrace and explain categorical-authoritativeness phenomenology, alongside the non-naturalist realist way and the error-theory way. That will suffice to refute Argument 2, leaving our version of expressivism in the running as a third option for coming to terms with the phenomenology. Wider, largely non-phenomenological, considerations thereafter can be brought to bear in doing overall cost-benefit evaluation of those three competing meta-ethical positions—a matter we will briefly address in Section 5 below. (We emphasize that we are not pre-supposing the Neutrality thesis at the outset. Rather—and granting, as we do, that the other two positions also can acknowledge and explain the perti-nent phenomenology—the case for Neutrality will emerge as the discussion below unfolds.)

In Subsections 4.1 and 4.2, we will make some preliminary method-ological observations. In Subsection 4.3, we will briefly sketch our favored version of expressivism, which is elaborated at greater length elsewhere (Horgan & Timmons 2006; 2015; 2017). In Subsection 4.4, we will set forth a concrete hypothetical moral-decision scenario for purposes of illus-tration of the points to come. With all this in place, in Subsection 4.5, we will turn directly to the business at hand.

4.1. *Expressivism About Reasons-Experiences and Reasons-Judgments*

Expressivists focus their meta-ethical theorizing, first and foremost, on the pertinent *states of mind* associated with matters ethical: moral experi-ences and moral judgments. Moral assertions are then treated as *expressing* moral judgments but not as *describing them*. A key tenet of any version of expressivism is that moral experiences and moral judgments do not pur-port to attribute in-the-world, instantiated, moral properties or relations, or to describe in-the-world moral facts. (This does not preclude the use of property-talk or relation-talk or fact-talk or truth-talk in moral discourse, but it does construe such talk as operating minimalistically rather than car-rying ontological purport.)

Typically, when one forms a moral judgment, one experiences certain considerations as *reasons* for that judgment—and indeed, as categorically authoritative reasons. Such reasons-experiences should be construed by

expressivists in the same kind of way that moral judgments themselves are construed—viz., as states of mind that do not purport to attribute in-the-world relations (say, *being fitting in light of*, *being a reason for*, or *being required by*, or the like). This point was emphasized by Charles Stevenson (1961), and we ourselves heartily concur.

So the task we face has two intertwined aspects. On one hand is the need to provide an intelligible and plausible expressivist construal of moral reasons-experiences themselves, over and above the moral judgments to which they give rise. On the other hand is the need to do so in a way that also accommodates the phenomenological aspect of categorically authoritative moral fittingness that Hampton so rightly emphasizes is central to moral experience.

4.2.   *Reductive vs. Non-Reductive Expressivism*

Both advocates and opponents of expressivism often characterize moral experiences and moral judgments by reference to states of mind which quite clearly do not purport to represent in-the-world instantiated properties or relations or in-the-world facts. Sometimes the reference to such states of mind is put forth merely as an instructive analogy; sometimes the suggestion is that although prototypical, such states of mind are not moral experiences, nevertheless the pertinent mental category includes moral experiences or moral judgments as a sub-species; and sometimes it is not made fully clear whether the first or the second construal is being suggested. At one time or another, expressivists have compared moral experiences and moral judgments to prototypical non-moral mental states of the following kinds, among others: desires, commands, universalized commands, attitudes of approval or disapproval, states of norm-acceptance, states of planning what to do.

In an important respect that is directly germane to the topic of the present essay, no such comparison—and no combination of such comparisons—is theoretically satisfying. The problem is this: for each such comparison-category, prototypical instances of that category are mental states that simply do not have the phenomenological aspect of categorical authoritativeness. Ordinary desires don't have it, because categorical authoritativeness is experienced as being independent of one's pre-existing desires. The mental states expressible as ordinary commands don't have it, because its phenomenology is not as-of a state of mind expressible by uttering a command, but rather (a) as-of being "commanded" oneself, and (b) as-of this command's having "compelling rightness." Ordinary states of norm-acceptance or action-planning don't have it, because typically they are experienced either as straightforwardly voluntary (as in voluntarily playing a game and thereby subjecting oneself to its rules), or at any rate as states one is in by virtue of one's contingent social circumstances (as in the case of the fictional character Ivan Denisovich (Solzhenitsyn 1962), who accepts the norms of bricklaying upon having been sentenced, in Stalin's post-war Soviet Union, to twenty-five years of bricklaying in Siberia). And so on.

One potential reaction to this problem facing such comparisons, of course, is to claim that the problem arises because moral experiences and moral judgments really do purport to represent in-the-world moral facts—including categorically authoritative, irreducibly normative, moral-fittingness facts.

But that is too quick. Instead, the appropriate reaction, we maintain, is to seek out an expressivistic construal of moral experiences and moral judgments that is *non-reductive*, in the sense that it treats these mental states as importantly different from paradigmatic mental states of those other kinds, and as *sui generis* in that respect—while nonetheless also treating them as not purporting to represent in-the-world instantiated moral properties, in-the-world instantiated moral relations, or in-the-world moral facts.

We ourselves have been pursuing this project in a number of prior writings (Horgan & Timmons 2006; 2008; 2015; 2017). We call our position *cognitivist expressivism*, because it treats moral judgment-states as a species of the genus *belief*—while yet also treating moral beliefs as importantly different from ordinary non-normative beliefs, and in this respect as being *sui generis*. We next summarize cognitivist expressivism as previously articulated. In Subsection 4.5, we will go on to elaborate it in a way that explicitly addresses the inherent phenomenological aspect of categorically authoritative fittingness in moral experience.

4.3.   *Cognitivist Expressivism*

Cognitivist expressivism is a meta-ethical position that is largely overlooked because of a widespread but (we claim) mistaken assumption—what we call the 'semantic assumption'. According to this assumption, every belief is to be understood as being a specific kind of commitment state—what we call an 'is' commitment—with respect to some *way-the-world-might-be* content (i.e., some *descriptive* content). For instance, one's non-moral belief (at some time t) that John took out the trash is to be understood as one's being is-committed (at t) with respect to a particular descriptive content—*that John took out the trash*. Likewise, if one's moral-judgment state expressible as "John ought to take out the trash" is a state of belief, then (given the semantic assumption) it is a state of being is-committed with respect to the putative descriptive content *that it ought to be that John takes out the trash*.

So in both non-moral and moral cases, according to the semantic assumption, to believe is to be is-committed vis-à-vis a way the world might be, i.e., vis-à-vis a descriptive content. Thus, if moral-judgment states are indeed beliefs, then they have *moral* descriptive content, i.e., they purport to represent instantiated in-the-world moral properties and relations.[3] So if one (*i*) embraces the semantic assumption, (*ii*) construes moral judgment-states as beliefs, and (*iii*) denies that there are any in-the-world instantiated moral properties or relations, then one is driven to a meta-ethical error theory.

Cognitivist expressivism, on the other hand, rejects the semantic assumption. This view posits two importantly different species of belief: in addition to is-commitments (call them is-beliefs), there are ought-commitments (ought-beliefs). The latter share much of the phenomenology of is-beliefs, and also share many key functional-role features possessed by is-beliefs.[4] (Possession of these common generic features is what qualifies ought-commitments as a species of the genus *belief*.) But an ought-belief involves a distinctive *kind* of commitment—an ought-commitment—directed toward a non-moral descriptive content. On this picture, to believe that John ought to take out the trash is to be *ought-committed* with respect to the non-moral descriptive way-the-world-might-be content *that John takes out the trash*. One might put the idea here by saying that on this conception of moral belief, the *ought* is in the attitude, rather than in a normative descriptive content toward which one is allegedly is-committed.[5]

Further delineating moral-judgment states as a distinctive species of the belief-genus is not a matter of trying to analogize them to, or subsume them under, states such as desires, commands, universalized commands, plans, norm-acceptances, or the like. This reductive approach, as already emphasized, looks incapable of accommodating the categorical-authoritativeness aspect of moral phenomenology. How then should one go about giving an illuminating expressivist characterization of these mental states? Well, in part by highlighting the generic features they have in common with mental states like non-normative beliefs, and in part by highlighting the *distinctive* features of their phenomenology together with the associated distinctive features of their functional roles in thought and in action-guidance. Our plan here is to do exactly that, in a way that turns Argument 2 from Subsection 3.3 on its head. There are two intertwined aspects of this strategy: first, embracing, as distinctive of moral experience and thought, the very categorical-authoritativeness phenomenology that Hampton stresses; and second, explaining why this phenomenology need not be construed as carrying objectivist ontological purport.

The following point deserves emphasis too. For present purposes, it will not matter whether or not we are right in claiming that ought-commitments are correctly construed as a species of belief. In principle, one could repudiate this specific claim while yet accepting everything else we say below. The important contention here is that moral judgments are psychological commitment-states of a certain distinctive kind: states which, *inter alia*, have the inherent phenomenological aspect of categorically authoritative moral fittingness, but which nonetheless do not presuppose any such putative in-the-world phenomenon as categorical moral authority and do not purport to represent in-the-world moral properties, relations, or states of affairs.

4.4. A Concrete Scenario: Clive's Callousness

As a prelude to the business at hand, it will be helpful to have in view a specific scenario involving an agent's morally tinged decision and its psychological aftermath. The (fictional) scenario is from Ian McEwan's (1998) novel *Amsterdam*, in which the character Clive Linley, a Londoner and a composer of some notoriety, is struggling to complete a symphony, celebrating the new millennium, for an upcoming concert in Amsterdam. His aim is to compose a masterpiece that will become the crowning achievement of his already illustrious career. Frustrated with a succession of failed attempts to compose the finale, Linley decides to seek inspiration by spending time away from the city, hiking in the Lake District located in a mountainous region of northwest England. While hiking, Linley is suddenly struck with an idea for the finale, and stops to scribble notes, attempting to work out the melody. His concentration is interrupted when he hears the nearby voices of a man and a woman quarreling. Clive climbs to the top of a large rock where he can see the quarreling couple standing face to face in a small clearing about thirty yards away. As the confrontation continues, loud talking soon gives way to shouting. The man grabs the woman's elbow, violently pulling her in his direction. Witnessing all this, Clive, with pencil and notebook in hand, sighs and ponders what to do:

> Was he really going to intervene? He imagined running down there. The point at which he reached them was when the possibilities would branch: the man might run off; the woman would be grateful. . . . Even this least probable of outcomes would destroy his fragile inspiration. The man was more likely to redirect his aggression at Clive while the woman looked on, helpless. Or gratified, for that was possible too; they might be closely bound, they might both turn on him for presuming to interfere. . . . What was clear now was the pressure of choice: he should either go down and protect the woman, if she needed protection, or he should creep away. . . . He could not remain here doing nothing.
>
> (1998: 93)

Clive closes his eyes and tries to concentrate on the elusive melody he is after. But at the sound of angry voices, he takes another look. The woman breaks loose of the man's hold with a sharp downward jerk of her arm, and turns to run, but the man tackles her from behind. They fall to the ground, the woman trying to crawl away, the man holding onto her ankles. The man, having gotten up, is now dragging the woman, both hands on her left ankle; she screams. Clive now understands the seriousness of the unfolding event; he thinks for a moment that he must intervene, come to the woman's aid. But the significance, the importance of his work! Clive hurries away from the scene, trying hard to recall those few notes of the melody he was so desperately trying to work out.

He was trying to call it back, but his concentration was being broken by another voice, the insistent, interior voice of self-justification: . . . if he had approached the couple, a pivotal moment in his career would have been destroyed.

(1998: 95)

Back from his hike, Clive decides to leave immediately for London, certain that he can work out the entire finale on the train. As he excitedly waits for his taxi ride to the train station, he reflects: "He wanted the anonymity of the city again, and the confinement of his studio, and—he had been thinking about this scrupulously—surely it was excitement that made him feel this way, not shame" (1998: 96).

For our purposes here, let us construe this scenario—by stipulation—as conforming to the following interpretation, which is suggested implicitly by McEwan's own characterization. Clive experiences being morally obligated to intervene in order to help the woman, despite deciding not to do so. The envisioned possibility of the man redirecting his aggression toward Clive, perhaps with the woman joining in, enters his mind primarily as a potential excuse for not intervening, whose flimsiness he already appreciates (he would be in no serious danger of significant bodily harm)—and which in any case, he realizes, is clearly outweighed by obligation once he sees the woman try to escape and the man then dragging her by the ankles. And the subsequent feeling that keeps him scrupulously thinking about what happened, rather than being able to concentrate on working out the finale of his symphony, really is shame rather than excitement.

4.5.   *Accommodating the Categorical-Authority Aspect of Moral Phenomenology*

We turn now to the task of offering, from within the framework of cognitivist expressivism, a non-reductive expressivist construal of the phenomenological aspect, in moral experience, of categorically authoritative fittingness. We will elaborate cognitivist expressivism beyond our earlier articulations of it, while also aiming to accommodate each of the elements of moral-authority phenomenology that Hampton stresses—the elements summarized at the end of Section 3 above. Indeed, we will underscore and seek to accommodate certain aspects of this phenomenology that Hampton herself does not stress, aspects that deserve acknowledgement regardless of one's preferred meta-ethical position. Focusing largely on the Clive example from Subsection 4.4, we aim to make clear why moral-reasons experiences can possess all the elements of moral-authority phenomenology that Hampton correctly attributes to them (and more besides) without purporting to represent in-the-world, categorically authoritative, fittingness-relations or fittingness-facts. The methodology will be *the phenomenological method of similarity and contrast* (as we will call it)—viz., comparing moral-reasons

experiences to various other kinds of mental states, noting both important phenomenological similarities and important phenomenological differences.

Consider the onset of Clive's non-normative belief that the woman is in serious danger of being harmed. He experiences what he sees happening—including, in particular, the woman's screaming while the man drags her by the ankles—as pulling him toward believing this—i.e., pulling toward an is-commitment vis-à-vis that content. (This is the aspect of *Pull*.) He experiences this as an "external" pull, emanating from what he sees happening outside of himself. (This is the aspect of *Source*.) He finds himself involuntarily gripped by an is-commitment vis-à-vis the likelihood of her being harmed, by virtue of the strength of this pull. (This is the aspect of *Grip*.) The onset of this is-commitment occurs independently of any pre-existing desires he has, and independently of human conventions—indeed, in this case it occurs despite his pre-existing desire *not* to believe that she is in serious danger. (This is the aspect of *Independence*.) Since his becoming thus is-committed is both involuntary and independent of his pre-existing desires, it is experienced as inescapable, given his current evidential situation. (This is the aspect of *Inescapability*.)

Consider now the onset of Clive's judgment that he is morally obligated to intervene on the woman's behalf—with this judgment being construed as an ought-commitment vis-à-vis the non-normative content *that I intervene on the woman's behalf*. This experience is phenomenologically similar, in each of the ways lately noted, to the onset of an is-commitment. He experiences her being in serious danger of being harmed (something about which he now has an is-commitment) as pulling him toward an ought-commitment with respect to his intervening. (This is the aspect of *Pull*.) He experiences this as an "external" pull, since her being in danger of harm is something outside of himself. (This is the aspect of *Source*.) He finds himself involuntarily gripped by an ought-commitment vis-à-vis intervening on her behalf, by virtue of the strength of this pull. (This is the aspect of *Grip*.) The onset of this ought-commitment occurs independently of any pre-existing desires he has, and independently of human conventions—indeed, in this case it occurs despite his pre-existing desire *not* to intervene. (This is the aspect of *Independence*.) Since his becoming thus ought-committed is both involuntary and independent of his pre-existing desires, it is experienced as *inescapable*, given his current evidential situation. (This is the aspect of *Inescapability*.)

The onset of Clive's is-commitment to the woman's being in danger of harm and the onset of his ought-commitment to intervening on her behalf are thus similar to one another in all the ways lately noted. (These similarities, we contend, are strong enough and broad enough to render the ought-commitment a species of the genus *belief* alongside the is-commitment—although our argumentation in this essay does not require agreement on this point.) In each case, the non-normative consideration experienced as *reason* for the commitment exhibits the same set

of elements that Hampton identifies as figuring in the phenomenology of categorical authoritativeness—viz., *Pull*, *Source*, *Grip*, *Independence*, and *Inescapability*.[6] And this is so even if—as we ourselves maintain—Clive's experiencing the woman's danger of harm as reason for his intervening does not purport to represent any putative in-the-world, putatively categorically authoritative, relation of "fittingness."

On the other hand, the onset of Clive's ought-commitment also differs in an important respect from the onset of his is-commitment regarding the woman's being in danger of harm—viz., the ought-commitment has a motivationally "hot" role within Clive's psychology, all by itself.[7] The is-commitment, by contrast—like other is-commitments to non-normative contents—plays no motivational role by itself apart from other pertinent psychological states with which it might combine, such as pre-existing desires. Thus, one's reason-experience in the case of the ought-commitment also has the final phenomenological element on the list at the end of Section 3, viz., *Motivation*. (This element can be present and operative without being motivationally dominant—as in Clive's case, since he does not act in accordance with his experienced moral obligation.)

The phenomenological element *Motivation* is also present, of course, in prototypical experiences of desire. In that respect at least, moral experiences are similar to ordinary desires—a point often emphasized by meta-ethical expressivists. Yet the differences from desire are palpable too. For one thing, desires are not commitment-states, whereas moral judgments are. And although a desire often is experienced as a reason for a specific action (given the belief that the action will lead to desire-satisfying consequences)—and thus exhibits *Pull* toward the action—the source of this pull is the desire itself, and so is not experientially external; thus, *Source* is not present. Also, acting on the basis of a desire, or forming an intention to act on the basis of a desire, typically is a voluntary matter, and thus does not exhibit *Inescapability*. And since desires themselves are the sources of the actions or intentions to which they sometimes give rise as reasons, such desire-based reasons do not exhibit *Independence*.

Yet further elements of the categorically-authoritative-fittingness aspect of moral phenomenology, beyond those explicitly mentioned by Hampton, can be brought into view by considering self-directed reactive attitudes that typically arise when one fails to act in a way that one experiences as morally obligatory—and by comparing these with self-directed reactive attitudes that typically arise when one violates an ongoing intention of a non-moral kind, such as an intention to stick to one's diet until one has lost ten pounds. In both kinds of case, one is apt to experience a sentiment of guilt or of shame, which in turn is apt to motivate one to take compensatory remedial action as best one can. And in both cases, such a sentiment is apt to arise because the pertinent state is experienced as exerting a governing authority over oneself—an authority that one has contravened. But there is a crucial difference. Although a voluntarily formed intention, such as the intention to

follow a specific diet because of one's desire to lose ten pounds, is apt to exert an experienced authority over oneself that will induce guilt and/or shame in circumstances where one has violated that intention, this authority and resultant sentiment(s) are experienced as *contingent and desire-dependent*: the authority one has contravened is experienced as operative upon oneself only in virtue of one's voluntarily formed intention, and thereby only in virtue of the pre-existing desire that motivated that intention in the first place. By contrast, the phenomenological authority of moral reasons, and the resultant self-directed reactive sentiment(s), are experienced as *inescapable and desire-independent*: the authority one has contravened is experienced as operative upon oneself independently of any medium-term or long-term intentions one might have voluntarily formed, and independently of one's pre-existing desires.

We submit that everything we have said in the present subsection about the inherent aspect of categorically authoritative fittingness in moral phenomenology is compatible with the contention that moral experience does not purport to represent putative in-the-world moral-fittingness facts. We are claiming no more than that, because as we emphasized earlier, the pertinent phenomenology also is compatible with the contention that moral experience *does* purport to represent in-the-world moral-fittingness facts. (We claim that direct introspection cannot reliably settle this issue either way, and that the phenomenology can be acknowledged by each of three competing meta-ethical positions: Hampton-style non-naturalist realism, Olson-style error theory, and our own non-reductive version of expressivism.) But the fact that our own expressivism can acknowledge and accommodate the categorical phenomenology of moral authority is all we need, for our dialectical purposes. Comparison of the respective costs and benefits of the three competing positions can now proceed in terms of extra-phenomenological considerations. We now turn briefly to that matter.

5.   Conclusion: Non-Reductive Expressivism vs. Non-Naturalist Realism and vs. Error Theory

In terms of comparative theoretical costs and benefits other than matters phenomenological, our non-reductive expressivism has all the same advantages over non-naturalist moral realism as does error theory. Most notably in the present dialectical context, of course, is the fact that both non-reductive expressivism and error theory avoid the extreme metaphysical queerness, from the perspective of a broadly naturalist metaphysical worldview, of putative in-the-world relations of categorically authoritative fittingness. Neither Mackie nor more recent error theorists could believe in in-the-world "to-be-done-ness" (as Mackie called it), and neither can we.

What about comparative theoretical costs and benefits of our version of expressivism vis-à-vis Olson-style error theory? Positing massive, population-wide, representational error in moral phenomenology is a

significant theoretical cost—especially if there is insufficient theoretical motivation for positing such error in the first place. Yet the fact that our brand of expressivism accommodates the categorical-authoritativeness aspect of moral phenomenology undermines the principal motivation for error theory, the motivation expressed in Argument 2 in Subsection 3.3 above. So our approach has a double theoretical advantage here: (*i*) undermining the main putative motivation for error theory, and (*ii*) avoiding the cost of attributing systematic error to moral experience—error which, if it were real, also would infect moral thought and moral discourse.

In addition, error theory faces a persistent theoretical problem anyway: what we will call the problem of *bad faith*. Error theorists usually are 'preservationists' rather than 'eliminativists' about moral thought and moral discourse: they advocate its continued use, even by those like themselves who believe that it, like moral phenomenology, purports to represent putative in-the-world, categorically authoritative, fittingness-relations that are in fact non-existent. (Mackie was a preservationist, and so are Olson and Joyce.) Preservationists recommend engaging in willful "as-if" moral thinking and moral discourse: in effect, *pretending* that there really are categorically authoritative in-the-world fittingness-facts—even when one confidently believes that this is not so. But it is very difficult to see how such an as-if stance can remain conceptually stable, rather than undermining itself in the minds of those who clear-headedly believe that moral experience, moral thought, and moral discourse are systematically non-veridical. This awkward conundrum is completely avoided by our own non-reductive expressivism.

Of course, an error theorist can avoid the bad-faith problem by going eliminativist rather than preservationist. This kind of error theorist proposes to maintain internal harmony between theory and practice by thoroughly revising practice itself: throwing out the baby (moral thought and discourse) with the bathwater (the metaphysical error putatively inherent to moral phenomenology and putatively present thereby in moral thought and moral discourse). But turning one's back on morals altogether is a theoretical last resort, surely to be avoided unless the reasons for doing so are enormously strong. And they aren't—because non-reductive expressivism accommodates, in a metaphysically irrealist manner, the aspect of categorically authoritative fittingness that is inherent to moral phenomenology.[8]

Notes

1 We are using the rubric 'moral experience' in a broad sense, to include not only making moral judgments but much more besides—for instance, feeling morally obligated, experiencing moral emotions (e.g., guilt, shame, and indignation), moral-perceptual experience, and experience as-of certain potential actions being morally fitting or morally unfitting under specific circumstances. These aspects of morality-involving mentality are conscious-as-opposed-to-unconscious, and in that sense are elements of moral *experience*. At a minimum, this means that they

are "access conscious" (in the influential terminology of Block 1995). In our own view, they also are *phenomenally* conscious: they have individuative phenomenal characters, and their phenomenal what-it-is-like-ness is the categorical psychological basis of their distinctive functional roles in human cognitive economy (including their availability for first-person report and the other functional roles characteristic of "access consciousness"). But those who have a less expansive view of the scope of phenomenal consciousness can, if they like, construe our use of the expression 'moral experience' in terms of mere access consciousness.

2 We understand this fittingness relation and the reasons relation to be the same relation. The former way of referring to this relation has as its focus an action or attitude being called for, and the latter way has as its focus the consideration doing the calling for.

3 Here we mean to be using the term 'descriptive', as applied to content talk, to include content involving putative irreducibly normative properties and relations.

4 We elaborate this claim in Horgan and Timmons (2006).

5 On our view, the *is* in the case of is-belief is also in the attitude. The idea is that the descriptive content of a belief—of either an is-belief or an ought-belief—is most perspicuously expressed not by a declarative sentence but rather by a 'that'-clause like "that John takes out the trash" (or by a nominalized sentence such as "John's taking out the trash"). In English, an is-commitment is canonically expressed linguistically by asserting a complete sentence in the declarative mood—as in "John took out the trash." An ought-commitment is canonically expressed by asserting a declarative-mood sentence whose predicative constituent comprises the modal auxiliary 'ought' appended to an infinitival verb—as in "John *ought to take out the trash.*" But on our account, the logical structure of is-beliefs and ought-beliefs is more perspicuously revealed via sentences employing a commitment-operator applied to a descriptive 'that'-clause, thus: "*It is the case* that John takes out the trash," "*It ought to be the case* that John takes out the trash."

6 Of course, an expressivist account like ours owes an expressivist-friendly explanation of these features. Here is not the place to delve into this; however, see our Horgan and Timmons (2015) for an explanation of how our view handles the issue of deep moral error, which bears on *Independence*.

7 This motivational role might be direct, or might be a matter of generating a new desire with the same content as the ought-commitment. We ourselves find the former possibility more plausible, phenomenologically and psychologically.

8 Thanks to David Copp, Diego Machuca, and Christine Tiefensee for their many helpful comments on an earlier version of this chapter.

References

Block, N. 1995. "On a Confusion about a Function of Consciousness," *Brain and Behavioral Sciences* 18: 227–247.

Enoch, D. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford & New York: Oxford University Press.

Garner, R. 1990. "On the Genuine Queerness of Moral Properties and Facts," *Australasian Journal of Philosophy* 68: 137–146.

Hampton, J. 1998. *The Authority of Reason*. Cambridge: Cambridge University Press.

Horgan, T., & Timmons, M. 2006. "Cognitivist Expressivism." In T. Horgan & M. Timmons (eds.), *Metaethics after Moore*, 255–298. Oxford & New York: Oxford University Press.

———. 2008. "What Can Moral Phenomenology Tell Us about Moral Objectivity?" *Social Philosophy & Policy* 25: 267–300.

———. 2015. "Modest Quasi-Realism and the Problem of Deep Moral Error." In R. N. Johnson & M. Smith (eds.), *Passions and Projections: Themes from the Philosophy of Simon Blackburn*, 190–209. Oxford & New York: Oxford University Press.

———. 2017. "Sentimentalist Moral-Perceptual Experience and Realist Pretensions: A Phenomenological Inquiry." In K. Stueber & R. Debes (eds.), *Moral Sentimentalism*, 86–106. Cambridge: Cambridge University Press.

Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.

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

McEwan, I. 1998. *Amsterdam*. New York: Anchor Books.

Nagel, T. 1986. *The View from Nowhere*. Oxford & New York: Oxford University Press.

Olson, J. 2014. *Moral Error Theory: History, Critique, Defense*. Oxford & New York: Oxford University Press.

Solzhenitsyn, A. 1962. *One Day in the Life of Ivan Denisovich*. Rutherford, NJ: Signet Classics.

Stevenson, C. L. 1961. "Relativism and Non-Relativism in the Theory of Value," *Proceedings and Addresses of the American Philosophical Association* 35: 25–44. (Reprinted as Chapter 5 of his *Fact and Value*. New Haven: Yale University Press, 1963).

7 Arguments From Moral Disagreement to Moral Skepticism

Richard Joyce

1. Introduction

References to the phenomenon of moral disagreement appear conspicuously in several areas in metaethics. One well-known argument—J. L. Mackie's (1977) *argument from relativity*—seeks to establish on the basis of widespread disagreement that moral judgments are false. Another argument—whose general form goes right back to classical philosophers like Agrippa and Sextus Empiricus—seeks to establish on the basis of widespread disagreement that moral judgments lack justification. Both conclusions can properly be called versions of moral skepticism. If moral skepticism is the view that there is no moral knowledge, then both the former error-theoretic position that moral judgments express false beliefs, and the latter epistemological position that moral judgments express unjustified beliefs, are forms of skepticism. (The noncognitivist position that moral judgments do not express beliefs at all is also a form of moral skepticism.)

This essay is almost exclusively diagnostic in ambition. I aim to sketch out the complex structure and interrelations of these skeptical arguments based on moral disagreement, without advocating any of them. Though I am very sympathetic to moral skepticism, I am yet to be convinced that there is an argument for it based on the phenomenon of moral disagreement that I would find persuasive in the absence of that sympathy. There are a couple of recurring themes that are worth highlighting in advance. One is the fact that the debate frequently hinges on empirical matters; the other is the complicated relation between skepticism and moral naturalism: often, as we shall see, the skeptic is opposed to naturalism, but other times the possibility of a defensible moral naturalism turns out to be a skeptic-friendly result.

2. Mackie's Argument From Disagreement

Mackie's argument from relativity is poorly named, since that title might be taken to suggest that he is seeking to establish moral relativism, which he

is definitely not. I prefer to call it "the argument from disagreement."[1] As a first stab at the argument, consider this:

Argument From Quantity of Disagreement

P1: If moral realism were true, then we would observe no more than such-and-such amount of moral disagreement in the world.
P2: In fact, we observe a great deal more moral disagreement than such-and-such.
C: Therefore, moral realism is false.

Even if we could correct for the glaring imprecision of "such-and-such," both premises are vulnerable to realist objections. Realists object to P2 by trying to downplay the amount of moral disagreement we actually observe. Much of what we might think of as moral disagreement, they say, is really disagreement over non-moral beliefs masking more fundamental moral agreement. And realists cast doubt on P1 by pointing out that it is not clear how much convergence or divergence in moral opinion moral realism really predicts. They seek partners in innocence: other domains where there is a great deal of disagreement (e.g., competing scientific hypotheses) but for which we are very disinclined to reach for an anti-realist conclusion.[2]

In response, anti-realists can point out that it's not just the *amount* of disagreement that's the issue; it's the *nature* of it. Moral disagreement, they might say, is characterized by an unusual kind of intractability, persistence, emotiveness, and insensitivity to evidence. In fact (they might add), it's often not even clear what would count as evidence for the truth of one moral judgment as opposed to a contrary one (Harman 1977). And perhaps the disagreement that attends the partners in innocence, widespread though it may be, lacks these qualities. So the argument shifts focus:

Argument From Quality of Disagreement

P1: If moral realism were true, then moral disagreement would not be so intractable, persistent, emotive, and insensitive to evidence (etc.).
P2: Moral disagreement is so intractable, persistent, emotive, and insensitive to evidence (etc.).
C: Therefore, moral realism is false.

Again, realists may object to P2 by claiming that there is more hope of convergence in moral disputes than Mackie suggests. If many moral disagreements are really at bottom disagreements over non-moral matters, then perhaps they are not so intractable after all. Realists can cast doubt on P1 by pointing out that it is not clear how much intractability, etc., of moral disagreement would be predicted by moral realism; perhaps moral realism is compatible with moral disagreement's having these qualities. At this point, Mackie deploys an argument with the form of inference to the best

explanation. The phenomenon to be explained is moral disagreement ("the well-known variation in moral codes from one society to another and from other period to another, and also the differences in moral beliefs between different groups and classes with a complex community" [1977: 36]), bearing in mind both its quantity and aforementioned qualities. Mackie compares two explanatory hypotheses. First there is the realist hypothesis: that moral codes "express perceptions, most of them seriously inadequate and badly distorted, of objective values" (1977: 37). As for the second hypothesis—the one which Mackie prefers—we shall have to have some discussion of its content. At the very least, it involves the claim that:

(*i*)   Moral codes "reflect people's adherence to and participation in different ways of life."

(1977: 36)

On this reading, though, it is difficult to see how an anti-realist conclusion is supposed to follow, even if we have grounds for thinking that the second hypothesis is correct, for the truth of (*i*) doesn't obviously exclude the truth of the realist hypothesis.[3] Indeed, (*i*) looks rather like a friendly supplement to the realist hypothesis: the realist hypothesis stating that many of our moral perceptions are "badly distorted," and (*i*) then explaining why this is so. In light of this, it is tempting to read the second hypothesis as requiring an additional claim:

(*ii*)   . . . and there are no objective moral facts.

On the realist's hypothesis, then, there are objective moral facts, but because we disagree so much we must not be very good at accessing those facts; and on the anti-realist's hypothesis, there are no objective moral facts and our disagreement is the result of our essentially "making up" morality to suit various practical needs that differ among cultures and individuals. And then Mackie invites us to agree with him that the second explanatory hypothesis is much more plausible.

One thing that should be noticed about this argument is that even if Mackie were to succeed in establishing the second hypothesis over the first, he still falls short of his ultimate metaethical conclusion: error theory. Error theory is not the view that there are no objective moral facts; it is the view that there are no moral facts *simpliciter*. Even if one agrees that there are no objective moral facts, one might well embrace a non-error-theoretic view that there are *non-objective* moral facts (e.g., some form of constructivism). To exclude this possibility, Mackie might countenance an even stronger version of the anti-realist hypothesis, one that combines (*i*) with:

(*iii*)   . . . and there are no moral facts.

The reason, I think, that Mackie is not too bothered in this context about the difference between (*ii*) and (*iii*) is that by the time that he puts forward

the argument from disagreement, he has already, to his own satisfaction, established that objectivity is an essential feature, conceptually speaking, of moral properties. Thus the difference between (*ii*) and (*iii*) is rather like the difference between "There are no four-sided squares in the box" and "There are no squares in the box." Since squares are necessarily four-sided, showing that there are no four-sided squares suffices to show that there are no squares *simpliciter*; likewise, if moral facts are necessarily objective, then showing that there are no objective moral facts suffices to show that there are no moral facts *simpliciter*.

One might point out that (*i*) + (*iii*) *still* doesn't entail the moral error theory, since it is compatible with noncognitivism. The error theorist and the noncognitivist agree that there are no moral facts; where they disagree is that the former maintains that moral speakers attempt to state moral facts, whereas the latter holds that moral discourse was never in the fact-stating business to begin with. But, again, by the time Mackie gets to presenting the argument from disagreement in his 1977 book, he has already put forward arguments against noncognitivism. Thus, by establishing the (*i*) + (*ii*) hypothesis over the realist hypothesis, Mackie thinks he is establishing the moral error theory, but only with the help of arguments that have come earlier in his essay.

The problem with adding (*ii*) or (*iii*) to the second hypothesis, however, is that doing so seriously undermines the grounds we might have for endorsing it. The claim made by (*i*) is at bottom an empirical claim. Cross-cultural investigation might reveal that societies do indeed construct their moral codes to suit their circumstances. It might reveal, for example, that societies that permit polyandrous marriage arrangements do so because of some unusual environmental feature, such as a paucity of farmable land (see Starkweather & Hames 2012). It might reveal that large-scale and complex societies are more likely than smaller and simpler societies to endorse and enforce fairness norms governing interactions with strangers (see Ensminger & Henrich 2014). And so on. In principle, then, we might muster evidence in support of (*i*), *but none of this evidence would support (ii) or (iii)*. This raises the question of why anyone, in attempting to explain moral disagreement, would prefer to maintain either of the hypotheses (*i*) + (*ii*) or (*i*) + (*iii*) over the less committed hypothesis (*i*). After all, the hypothesis that denies objective moral facts or denies moral facts will be worse off than the agnostic hypothesis if it turns out that there are phenomena other than moral disagreement whose explanations *do* require the existence of these facts (a possibility that, if we are seeking only to explain moral disagreement, we have no grounds for excluding). Moreover, the additional clauses (*ii*) or (*iii*) won't serve to better explain any aspect of the phenomenon of disagreement. In fact, what business does an explanation have in denying the existence of something? Compare the perfectly reasonable claim "The best explanation of moral disagreement remains silent on the existence of wombats" to the bizarre claim "The best explanation of moral disagreement denies the existence of wombats." (Note that the contrast remains if

we replace mention of wombats with mention of something in which we don't believe—unicorns, say.)

It becomes clear, then, that the active denials embodied in (*ii*) and (*iii*) should not be considered elements of the explanatory hypothesis. The explanatory hypothesis is simply proposition (*i*)—though of course (*i*) as it is worded here is merely a stub that stands in for a much more complex account—which is silent on the existence of both objective moral facts and moral facts *simpliciter*. But the problem remains that even if (*i*) is the best explanation of the moral disagreement, it is entirely compatible with moral realism. We might at this point declare the argument from disagreement a flop. Or we might try to salvage a case for the error-theoretic conclusion from the pieces of the argument that are lying on the table. Let's try the latter.

3.   Arguing for Error-Theoretic Skepticism

Mackie evidently needs to combine proposition (*i*) with some other premise(s) in order to produce an error-theoretic conclusion. What would serve as a bridging premise is some kind of principle of parsimony that allows the transition from judging that moral facts (or just objective moral facts) have no place in an explanation to concluding that they don't exist. Both ends of this transition need more scrutiny.

First, showing that moral facts (or objective moral facts) play no role in the explanation of some limited phenomenon, such as disagreement, is surely not going to warrant the wholesale denial of their existence, for (as noted above) moral facts (or objective moral facts) may be needed to explain some other phenomenon.[4] If Mackie is going to have any hope of getting to such a grand anti-realist conclusion, then he must be confident not only that moral facts play no role in explaining moral disagreement, but that they play no role in the explanation of any phenomenon at all. If we are considering *everything* that we know of, then the anti-realist hypothesis is not vulnerable to being overturned by consideration of some other phenomenon that requires a realist explanation, for *ex hypothesi* we know of no such phenomenon.

Second, even it is true that there is no phenomenon whose explanation requires the positing of moral facts, one might still wonder on what grounds someone would prefer denial to agnosticism. The crucial difference is between the following two bridging principles of parsimony:

(A) If something plays no explanatory role, then we have no ground for believing in it.
(B) If something plays no explanatory role, then we have ground for disbelieving in it.

A philosopher may of course simply announce that (B) is one of his or her basic methodological principles. The logical positivists' enthusiasm for the

verification principle smacked of this kind of programmatic decree (to say nothing of Hume's rhetorical advice that anything not meeting his empiricist standards should be "committed to the flames"). But simply claiming (B) to be methodological bedrock may seem like a dogmatic overreach when (A) is also available, for surely (A) is more reasonable. Certainly, if we are considering the explanation of a limited phenomenon, then this seems to be the case. The fact that unicorns play no role in explaining X (choose any ordinary phenomenon here) may well provide us with no ground for believing in unicorns, but nor does it provide grounds for actively *disbelieving* in them. But if we lift the limits and consider *all* known phenomena, then there appears to be more to be said in favor of the stronger principle (B). The fact that unicorns play no role in *any* good explanation—that, in other words, we have no evidence for their existence whatsoever—may, one might think, be precisely why it is reasonable to *disbelieve* in them.

Things are somewhat more complicated than this, however. Compare this with a case for which agnosticism intuitively seems the correct epistemic attitude. Are there planets orbiting the star Betelgeuse? There's currently no evidence one way or the other (so far as I can tell from a quick google search). On the basis of current evidence, one should neither believe that there are planets orbiting Betelgeuse, nor disbelieve this. The presence of planets orbiting Betelgeuse plays no role in explaining any phenomenon we know of, and yet this doesn't seem to provide grounds for disbelief.

The difference between the two cases is that given the obvious current limitations in our ability to gather evidence about what's going on in distant solar systems, it is no surprise that we have no evidence one way or the other regarding Betelgeuse's planets. That we should have any evidence at all about planets orbiting distant stars is still a relatively novel idea; we do not expect to currently have evidence one way or the other regarding a great many stars. By comparison, in the case of unicorns we do expect that if they existed anywhere (on Earth), by this stage we'd have uncovered some evidence of the fact. Perhaps there was a time in the Middle Ages when agnosticism about unicorns was appropriate, but as we explored more and more of the world and uncovered no evidence, the reasonableness of agnosticism gave way to the reasonableness of disbelief. Thus we see that (B) is plausible only with amendment:

> (B*) If something plays no explanatory role, then we have ground for disbelieving in it, if it is reasonable to assume that if it existed then we would have evidence of it.[5]

So are moral facts more like unicorns or like planets orbiting a distant star?

Answering that moral facts are (in this respect) more like unicorns seems a perfectly coherent thing to say. There is little doubt that this is what Mackie would say, since the claim that if moral realism were true then we would expect to have evidence of moral facts is very close to the claim he

makes in the argument from disagreement: that if moral realism were true, then we would expect to observe fewer moral disagreements. However, it is important to see that plausible options remain open to the realist to resist the skeptical conclusion.

First, of course, moral realists need not accept the claim that moral facts aren't required to explain anything—they can maintain that moral facts play all sorts of roles in explanations. Nicholas Sturgeon (1985) famously considers the case of Selim Woodworth, who in 1846 contributed substantially to the Donner party's unhappy fate through his ineffective and incompetent leadership of a rescue effort. The historian Bernard deVoto concluded that Woodworth was "no damned good," and Sturgeon claims that the best explanation of deVoto's forming this belief is that Woodworth was, in fact, no damned good. Sturgeon also cites the example of Hitler's moral depravity as the explanation for his ordering the death of millions of people.

One might object to the realist's argument by claiming that whenever we have a moral explanation for a phenomenon, there is always a superior non-moral explanation available. After all, we do not need to refer to Woodworth's being "no damned good" in order to explain his decisions; we could refer instead to psychological factors like his ambition, his lack of empathy, and so forth. In turn, we can explain deVoto's forming the judgment that Woodworth was no damned good by reference to his having certain beliefs about how Woodworth acted, coupled with deVoto's commitment to certain moral values which he had come to internalize through a (no doubt complicated) process of socialization. Had deVoto been raised differently, perhaps he would not have condemned Woodworth's actions in this manner.

But why is the non-moral explanation superior? Why, in fact, is it a competing explanation at all? The anti-realist might try to answer the first question by appeal to parsimony. The non-moral explanation is preferable because it posits less—in particular, it doesn't require the existence of moral facts. This answer, however, presupposes that the moral facts in question must be something "extra" in an ontological sense, and this is something that a naturalistic moral realist simply denies. The naturalistic moral realist identifies moral properties with naturalistic properties that are already present in the ontological frameworks accepted by all parties involved. If, for example, we can explain Hitler's actions either by reference to his upbringing, situation, and personality traits (i.e., in non-moral terms) or by reference to his depravity (i.e., in moral terms), then the latter is no more ontologically extravagant than the former if the property of being morally depraved just *is* the having of those personality traits. By analogy, if we face a choice of explaining a phenomenon (e.g., rust on the exhaust pipe) either by reference to the presence of pairs of hydrogen atoms bonded with single oxygen atoms, or by reference to the presence of water, then neither explanation is ontologically cheaper or costlier than the other. In fact, although one explanation may be more pragmatically suitable than the other to certain conversational contexts, they are not really competing explanations at all.[6]

Since we are trying to understand how Mackie might make his argument from disagreement work, it is worth noting that the aforementioned defensive strategy from the moral realist will not move him. Earlier I pointed out that it is a mistake to read the two pages of Mackie's argument from disagreement in isolation and expect to find a persuasive argument contained therein, for important premises (rejecting noncognitivism, rejecting any non-objectivist construal of moral facts) have already been argued for in preceding pages. (In saying this I am not claiming, of course, that Mackie's earlier arguments are entirely *convincing*—far from it—I'm just trying to straighten out the moves.) We find the same pattern here. Prior to offering his argument from disagreement, Mackie has already, to his own satisfaction, deployed considerations against the moral naturalist:

> On a naturalist analysis, moral judgements can be practical, but their practicality is wholly relative to desires or possible satisfactions of the person or persons whose actions are to be guided; but moral judgements seem to say more than this.
>
> (1977: 33)

Mackie has, in effect, argued that the only avenue available to the moral realist is a Moorean one, according to which moral properties really are non-naturalistic ontological "extras." This is why the obvious realist move of claiming that moral properties do have an explanatory role, in virtue of their being identical to (or supervening on) explanatorily potent non-moral objective properties, won't cut any ice with Mackie. Because he sees the moral realist as having non-naturalist commitments, he thinks that there are grounds for claiming that whenever we are presented with a moral explanation for a phenomenon, there will always be a superior non-moral explanation available.

The second thing that the moral realist can say against (B*)'s applying to moral facts is that for certain moral facts, it is not reasonable to assume that we should have evidence for them (yet). (The corollary of this argument is that it is not reasonable to assume that if there are objective moral facts, then there would be less moral disagreement than there actually is.) Derek Parfit (1984: 454) observes that secular ethics is a young discipline, and thus our evidence-gathering methods remain immature. And even when they mature, there may be no guarantee of complete convergence. As David Brink writes: "Moral ties are possible, and considerations, each of which is objectively valuable, may be incommensurable" (Brink 1984: 116; see also Shafer-Landau 1994; Harman & Thomson 1996: 205–206). In such cases, even though there may be moral facts, we would not have reliable evidence one way or the other about them; the evidence would be permanently unclear, and thus disagreement would persist. Some moral facts may simply be enormously difficult to apprehend. While we can be reasonably confident that in principle we could find out whether there are planets orbiting

Betelgeuse, we also recognize that doing so currently surpasses our epistemic abilities. Similarly, some moral facts might depend on (for example) a delicate balance of future painful and pleasurable consequences, the knowledge of which we might suppose that we could *in principle* gain, but for which we recognize that doing so currently surpasses our epistemic abilities. It is worth remembering that realism is, strictly speaking, entirely compatible with radical skepticism, according to which moral knowledge is *impossible*. Of course, the realist is unlikely to be attracted to that position,[7] but it remains an open question what the realist might say about how easily that knowledge comes. Thus the realist might well claim that moral facts are, in this crucial respect, like the planets orbiting a distant star: the fact that we need not posit them to explain any phenomenon doesn't count against their existence, since it is no surprise that we currently lack evidence.

The discussion thus far has involved a lot of moves and counter-moves, so it may be worth pausing to take stock. We started out wondering how Mackie's argument from disagreement is supposed to yield his preferred error-theoretic conclusion. The initial question of whether an anti-realist explanation of moral disagreement is superior to a realist explanation of moral disagreement proved to be problematic, since even if the answer were "yes," we would be none the wiser as to whether objective moral facts might be needed to explain some other phenomenon. We were forced to step back and ask the broader question of whether moral facts are required to explain *anything*. Our attention alighted on principle (B), which was then amended to (B*). The issue isn't whether (B*) is true—let's assume it is—the issue is whether the principle applies to moral facts. To assess this matter, two questions must be scrutinized. The first is whether it is true that moral facts play no explanatory role. I pointed out that Mackie's positive answer to this question depends on his defeating the possibility of moral naturalism—the arguments for which he presents prior to offering the argument from disagreement. (I have not, however, tried to evaluate Mackie's argument against moral naturalism, though I happen to think he's right [see Joyce 2001; 2006: ch. 6; 2016a: 380–381].) The second question is whether it is reasonable to assume that if there are moral facts we would likely have evidence of them. (This is really just a more general way of asking the question posed by the argument from disagreement: if there are moral facts, then shouldn't there be less moral disagreement than there actually is?) This is a very tricky question to address; it depends very much on the realist's particular conception of the nature of moral facts.

So we haven't gotten very far in establishing whether Mackie's argument from disagreement can be developed into a sound basis for the moral error theory, but we have identified where the battle lines might be drawn, and at least established the perhaps disappointingly exegetical conclusion that while the argument from disagreement gives the illusion of focus— it is, after all, only two pages long—what is really powering it are much larger issues (concerning not just whether moral facts are needed to explain

disagreement, but whether they are needed to explain anything) and arguments that lie elsewhere.

4.   Arguing for Justification Skepticism via Genealogical Debunking

Rather than wondering how Mackie might press this argument for error theory more effectively, now I want to consider the possibility of his backing off from that strong conclusion and aiming for a less ambitious form of skepticism (not all-things-considered, but just with respect to the argument from disagreement). Suppose he were to accept (A) instead of (B*):

(A) If something plays no explanatory role, then we have no ground for believing in it.

Because (A) is more modest than (B), there is no need to amend it in the analogous way that (B) became (B*). And because these are bridging principles, the conclusion at the far end of the bridge can be proportionally weaker. We are no longer aiming for (*ii*) or (*iii*), but their less presumptuous counterparts:

(*ii**) . . . and there may be no objective moral facts.
(*iii**) . . . and there may be no moral facts.

The anti-realist still has to establish that moral facts (or objective moral facts) play no explanatory role, and we know that the realist need not concede that point without a serious fight. Most of what I have already said about that debate holds as much for this weaker argument as the previous stronger one. But the anti-realist has one less task to do: he or she doesn't have to argue for the difficult claim that it is reasonable to believe that if objective moral facts existed, then we would have evidence of them. Recall again the Betelgeuse case: I noted that the presence of planets orbiting the star plays no role in explaining any phenomenon we know of, and yet this doesn't seem to provide grounds for disbelief, since currently we wouldn't really expect to have evidence one way or the other. However, it's still reasonable to declare that because the presence of planets orbiting Betelgeuse plays no role in explaining any phenomenon we know of, *for all we know there may be no such planets*. In other words, though we might not have grounds for disbelief, nor do we have grounds for belief.

Such a position still counts as *skeptical*, though it is not the skepticism of the error theorist. Rather, just as we would say that someone who currently believes that there *are* planets orbiting Betelgeuse lacks justification for this belief and therefore lacks knowledge, so too (according to the view under discussion) someone who has any moral belief lacks justification and therefore lacks knowledge. Calling this conclusion "weaker" than the

error-theoretic result shouldn't lull us into failing to notice that it is still a radically skeptical view.

The argument under consideration is essentially a genealogical debunking one. Start again with hypothesis (*i*) (moral codes "reflect people's adherence to and participation in different ways of life")—an explanation that appears not to require the existence of moral facts. Again: the question of how to explain moral disagreement is just the point of departure; what really matters is whether moral facts are needed to explain *anything*. But as soon as we make the issue broader in this way, we should see that we can immediately shrink it again to the question of whether moral facts are needed to explain *moral judgments*. Let me explain. There is no phenomenon that we might be tempted to explain by reference to moral facts without our having made a moral judgment being an indispensable element. Consider Sturgeon's example of Hitler's actions, for instance, which he wants to explain by reference to Hitler's depravity. But one will be tempted by that explanation only if one is willing to judge Hitler's character depraved in the first place, and then the question becomes what explains *that* judgment: must we invoke Hitler's depravity to explain why someone judges him depraved, or can that judgment be explained better by a hypothesis that refers only to non-moral phenomena? If any person's moral judgment about Hitler can always be best explained without reference to moral facts, then we might conclude that those judgments lack justification. And if one's judgment that Hitler was depraved lacks justification, then no explanation that appeals to his depravity as an explanans (to explain genocide, etc.) should be accepted. In other words, the question of whether moral facts need be invoked to explain any phenomenon always boils down to the question of whether they need be invoked to explain moral judgment.

There has been quite a bit of discussion in recent years about genealogical debunking arguments (see Joyce 2006; 2016b; 2016c; Wielenberg 2010; Kahane 2011; Fraser 2014; Braddock 2016). The discussion has often taken the evolutionary perspective: arguing whether human moral thinking is the product of natural selection and, if so, what its adaptive purpose might have been. The evolutionary debunking argument has promise because it's reasonably plausible both to claim that moral thinking *is* the product of natural selection, and to claim that its evolutionary purpose was to play a role in strengthening our ancestors' social bonds so as to encourage them to cooperate together more effectively. The significance of the second claim is that moral thinking is explained in a way that makes no reference to any moral judgments being *true*. In this it contrasts with evolutionary accounts that might be given of other pieces of human psychology, such as our ability to recognize faces, for example. The evidence seems to indicate that humans have an innate mechanism for visually distinguishing faces from other stimuli (see Slater & Quinn 2001). In explaining why such a mechanism might have evolved, one is likely to mention the importance of social bonding early in infancy, the stability of the presence and anatomical

structure of human faces, and so on. The crucial feature of this explanation, though, is that it presupposes that faces actually existed in the ancestral environment—the face-identifying mechanism was useful precisely because it reliably succeeded in putting the infant in causal contact with actual faces. The evolutionary account of the human capacity to make moral judgments differs critically in this respect. The explanation is (very roughly) that having the ability to judge certain actions as morally required (say) was useful because it encouraged our ancestors to cooperate in fitness-enhancing ways—*not* because it allowed them to identify which actions really did have the property of being morally required. The view is not that particular moral judgments are hard-wired; it is that the basic capacity to employ a moral conceptual framework is hard-wired, and then the social environment determines which moral norms one ends up endorsing. (In an analogous way, humans may have evolved mechanisms dedicated to acquiring a language, but precisely which language a person ends up speaking is determined by the social environment.)

Despite the focus on the evolutionary perspective in recent literature, a genealogical debunking argument can run without it. What matters is that a plausible, or perhaps even empirically confirmed, complete account of moral judgment can be given which does not imply or presuppose that moral judgments are true or even probably true. Instead of the explanation being at the evolutionary level, it might instead be at the anthropological level (how cultures come to adopt their moral norms) or at the level of developmental psychology (how individuals come to internalize moral norms).

But any genealogical argument is susceptible to the same objection that we encountered earlier: even if moral judgments can be explained entirely in non-moral terms, this does not show that moral facts are explanatorily impotent *if moral facts are identical to those non-moral facts mentioned in the explanation*. Forms of moral naturalism promise to establish just such an identity relation.[8]

One strategy for the proponent of the debunking argument, then, would be to supplement it with anti-naturalist arguments. (These arguments would just be against *moral* naturalism, of course; they could be entirely consistent with a more general commitment to methodological naturalism.) This is the strategy I pursued on an earlier occasion (Joyce 2006), peddling arguments that can be seen as elaborations of the same doubts that Mackie voiced against moral naturalism, quoted briefly earlier. In fact, however, such arguments might be considered overkill. All that really needs to be established is that particular versions of moral naturalism are unacceptable—namely, those that would relate moral facts to those naturalistic properties mentioned in the genealogical explanation. Thus the proponent of the debunking argument need not have a prior commitment opposed to the very idea of moral naturalism (*contra* Das 2016).

Another strategy for the proponent of the debunking argument (and, for what it's worth, one that I've preferred in recent years) is to see the argument

in terms of establishing a burden of proof. It's not enough to say that there *might* be identity relations holding between moral properties and those naturalistic properties explicitly cited in the genealogical explanation—rather, the proof of the pudding is in the eating. After all, even the moral error theorist is going to agree that moral judgments have some kind of history—whether at a psychological, anthropological, or evolutionary level—and someone can always claim that it's *possible* that the moral facts are surreptitiously buried somewhere among the historical facts that the error theorist is willing to accept. However, it seems reasonable for the error theorist to remain unbothered by this claim until the naturalistic account is displayed and defended. The conclusion is not that we should provisionally assume that the moral error theorist is correct in thinking that moral judgments are false; the conclusion is that since moral judgments are the product of a process that appears to be consistent with a moral error theory, then we should provisionally assume that they lack justification.

Consider an analogy. Suppose it turns out that when asked about the capital of Gabon, I find myself spontaneously answering "Freetown." I assume it's something I picked up somewhere, though I can't recall any details. But now you show me some credible video footage from last Monday of someone asking me the capital of Gabon while I shrug and answer that I have no idea. Mysterious. Then you show me credible footage from last Tuesday of my being hypnotized by Madame K., who is telling me "When you awake, you shall believe that the capital of Gabon is Freetown, and you shall forget all about being hypnotized." Mystery solved. But now what attitude should I take toward the proposition that Freetown is the capital of Gabon? I could just look it up, but let's say that for some reason I haven't yet had the opportunity. It's clear that I really don't know what to think anymore. Maybe Freetown is the capital of Gabon, maybe it isn't. It would be reasonable to conclude that my previous, fairly confident belief lacked justification (or, at least, that if I persisted with the belief, it would now lack justification). It lacks justification because I've discovered that it's the result of a process that is consistent with its not being the case that Freetown is the capital of Gabon and does not render that fact very probable.

Perhaps I ask my friend Mary about the capital of Gabon, and am relieved to hear her confidently claim that it is Freetown. Justification reinstated! But then Mary and I are shown footage of her also being hypnotized by Madame K. last week ("The capital of Gabon is Freetown . . . the capital of Gabon is Freetown"), so now Mary and I are in the same confused epistemic state. And so we should be, for we have learned that our belief has come about through a process that disconnects it from the relevant facts. Mary has an idea: "Perhaps Madame K. hypnotizes people to believe *only true things* about countries' capitals!" This idea amounts to suggesting that the process that produced our belief *does* connect to the relevant facts (concerning African countries and cities), though somewhat more indirectly than we'd previously thought. But the mere possibility of this connection—of

Madame K.'s being epistemically benevolent in this manner—is not suffi-cient to reinstate justification. For that, we'd need some credible evidence that Madame K. actually does behave in this manner. In the same way, the mere possibility of a naturalistic theory connecting moral facts to the non-moral facts that figure in the error theorist's genealogy of moral judgment is not sufficient to reinstate justification. For that, the naturalistic theory needs to be made credible.

Whichever strategy the debunking skeptic prefers—whether going on the offensive and trying to refute versions of moral naturalism (or moral natu-ralism *simpliciter*), or defensively claiming that it is up to the naturalist to put forward a credible theory—it is clear that the debunking argument isn't designed to be an argument that defeats moral naturalism, but rather one that requires supplementation with anti-naturalist considerations. Recall that the same thing goes for the stronger error-theoretical skeptical argument that I examined earlier—the one revolving around (B*). This argument would also fail if certain versions of moral naturalism could be made plausible. It is also clear that these complex skeptical arguments have brought us a long way from the original presentation of Mackie's argument from disagreement.[9]

5.   Arguing for Justification Skepticism via Disagreement Among Peers

There is a more direct route from moral disagreement to justification skep-ticism. Proponents of the so-called *conciliatory view* on the epistemic sig-nificance of disagreement hold that when one encounters a disagreeing epistemic peer, one's epistemic confidence in the disputed claim should diminish (see Feldman 2006; Christensen 2007; Kornblith 2010; Matheson 2015).[10] The basic idea is quite mundane: suppose I glance at my watch and it reads 11:15; I therefore confidently believe that the time is 11:15, but then I notice that the clock on the wall reads 11:35. If I lack any grounds for privileging my watch over the clock, the fact that the clock "disagrees" with my watch should immediately give me pause. The confidence that I had in the belief that the time is 11:15 should be reduced; if I care about know-ing the time, then I'll need to take some steps to sort out the discrepancy. In an analogous manner, if you and I are splitting the tip at a restaurant, and I come to the result of paying $11.15 each but then you come to the result of $11.35 each, and I am as confident in your arithmetical abilities and honesty as I am in my own, then we have a puzzle: I should reduce my initial confidence in $11.15 as the answer and recalculate the sum more carefully. In an analogous manner again (supposedly), if I'm confident in judging that x is morally wrong, but I then encounter someone who thinks that x is morally acceptable, and I have no ground for privileging my own moral judgment-formation processes over those of the dissenter (i.e., I must accept that he or she is an epistemic peer in this matter), then my confidence in judging that x is morally wrong should be reduced.

Some have thought that disagreement with epistemic peers is so ubiquitous that a completely global skepticism follows. This was, famously, an important kind of argument put forward by classical skeptics of the Greek and Roman philosophy worlds.[11] Here we are interested in a more limited but still remarkable result: that there is something about moral disagreement in particular that leads, via a version of this argument, to moral skepticism. I have already remarked (regarding the argument from quality of disagreement) that moral disagreement seems to have qualities that are less characteristic of many non-moral disputes (intractability, etc.); perhaps this provides the basis of rendering moral skepticism plausible while allowing us to avoid the extravagant pessimism of global skepticism. As quick evidence of the difference, consider again the case of you and me coming up with different numbers when trying to split the restaurant tip. As mentioned, we would probably proceed by recalculating the sum more carefully. But suppose we both do so and I again come up with $11.15 and you again come up with $11.35. We frown and try again and the same thing happens. At this point we'd just be utterly baffled; it's not clear what we should do. But in the case of moral disagreement, we're reasonably tolerant of the possibility that no matter how much you and I deliberate carefully and "compare notes," I may simply continue to find *x* morally wrong while you continue to find it morally acceptable. We're not surprised that moral disagreements can persist in this manner; it's not baffling.

Central to the conciliatory view is the idea that one must be able to identify epistemic peers in a reasonable manner: those whose intelligence, freedom from bias, reflective awareness, access to and appreciation of the evidence, etc., are equal to one's own. One must, in short, be able to form justified views about others' epistemic credentials. Most proponents of the conciliatory view add some version of the *independence principle*: that in identifying epistemic peers, one must discount the fact of the dispute in question—you cannot, in other words, take the very fact that the person disagrees with you as evidence of her epistemic inferiority. If the dispute is over whether *p* is the case, then in evaluating your opponent's epistemic credentials you must ignore your belief that *p* and the reasoning that led you to that belief (and ignore her disbelief that *p* and the reasoning that led her to that disbelief). This principle promises to block the slide to wholesale skepticism. If a person disagrees with you about something incredibly fundamental, like whether the material world exists, then if, in assessing whether he is an epistemic peer, you must discount this belief of his and all the reasoning that led him to it, then it is unlikely that you will be left with sufficient resources to make a judgment of his epistemic credentials, in which case conciliationism simply remains silent on whether you should revise your belief on the matter (see Elga 2007; Vavova 2014).

Let us assume for the sake of argument that the conciliatory view is basically correct. Would moral skepticism follow? I will restrict myself to commenting on two reasons for thinking that it would not.

First, how one goes about assessing whether a moral disputant is an epistemic peer seems dependent on one's standing general attitude to the status of moral facts and evidence. Suppose Mary has already taken on board some of the worries that have been canvassed earlier in this essay: that moral facts do not seem to play a role in explaining any phenomenon, that it's not clear what would count as evidence for the truth of one moral judgment over another, that moral judgments can be explained in a way that appears consistent with an error theory, etc. Despite harboring these worries, Mary hasn't endorsed moral skepticism and continues to make moral judgments: she judges that *x* is morally wrong. She encounters Fred, who disagrees. Mary must now assess whether Fred is her epistemic peer. She can presumably wonder whether he is an epistemic peer on non-moral matters, but what about on moral matters? (It would seem strange to think that all she need concern herself with is whether Fred is an epistemic peer on *non-moral* matters. That would be like claiming that when Mary wonders whether Fred's disagreement over a math problem undermines her belief on that point, it suffices for her to ascertain whether he is her epistemic peer concerning horticulture.) While Mary can consider whether Fred is as intelligent and reflective as she is, how shall she assess whether Fred has equal access to and appreciation of the moral evidence? The natural concerns that non-moral disagreement might prompt—such as "Maybe he has better access to the evidence than I do" or "Maybe he appreciates the evidence better than I do"—may seem to Mary to be not even sensible worries to voice about moral beliefs. She's just not sure, in other words, whether, for moral cases, it makes sense to ask whether Fred is an *epistemic* peer. And if she cannot assess whether Fred is an epistemic peer, then the conciliatory argument remains silent on whether she should alter her confidence in the moral claim that *x* is wrong.

What is interesting in these thoughts is that Mary is sympathetic to views that (as we have seen earlier in this essay) can be used as premises in arguments for moral skepticism, but here those very same views are blocking the skeptical result. And the role of moral naturalism in the argument is reversed as well. In earlier arguments, a viable version of moral naturalism promised to thwart the skeptical argument. It is therefore somewhat ironic that here, if a viable version of moral naturalism were forthcoming, then Mary's worries would be answered, and she *would* be able to assess Fred as an epistemic peer (*ceteris paribus*), in which case the possibility that Fred's disagreement should lead Mary to downgrade her confidence in her moral judgment (i.e., the moral skeptic-friendly possibility) would reopen.

A second reason for thinking that moral skepticism does not follow from conciliationism follows from the "discounting" of beliefs required by the independence principle. Moral beliefs, the thought goes, are unlikely to be held in isolation from each other. If Mary and Fred's disagreement about whether *x* is morally wrong is quite fundamental, and if, in assessing whether Fred is an epistemic peer, Mary must discount her belief that *x* is morally wrong *and all*

the reasoning and evidence that led her to that belief, then her resources for deciding whether Fred is an epistemic peer diminish. Katia Vavova pushes this argument, maintaining that "as our disagreement deepens, the grounds I have for taking you to be my peer shrink" (2014: 314). If, on the other hand, Mary and Fred's moral disagreement is relatively shallow—if, that is, they have a track record of agreeing on most moral matters, but have encountered this one difference in moral opinion—then an epistemic retreat from confidence to agnosticism, as conciliationism appears to demand, would be entirely appropriate. But since (Vavova thinks) such appropriate reductions in moral confidence in response to disagreement are sufficiently rare and reasonable, then they represent no comment-worthy concession to moral skepticism.

It seems to me that this argument is a hostage to empirical fortune. It is far from clear to what extent people's moral beliefs are based on "reason and evidence," and the assumption that people derive their moral judgments from broader moral principles, which in turn are based on the endorsement of fundamental moral values, may be something of a moral philosopher's optimistic projection of *Ethics 101* onto human moral psychology. While it seems true that moral judgments often come in "packages" (e.g., in the US, someone's views on abortion likely correlate with their views on gun control and taxation), it doesn't follow that this is because these views have been inferred from deeper moral principles. It may be, rather, that certain personality traits, like risk-aversion and disgust-sensitivity, lie behind these judgments (see Choma *et al.* 2014; Hibbing *et al.* 2014), and that lying behind some of these differences in personality traits is genetic variation (see Benjamin *et al.* 2012; Hatemi & McDermott 2012). I am simplifying things exceedingly in order to make the point succinctly—the point being that there may be *causal* connections among packages of moral judgments (e.g., being jointly caused by some neurological trait), but the demand made by the independence principle—that one must ignore the disputed belief and all the reasons that have led one to it—concerns the inferential process that has led one to the belief, not the belief's psychological causal antecedents.

Vavova's view is that in deep moral disputes, the independence principle requires one to bracket off so much that one no longer has grounds to judge whether one's disputant is an epistemic peer. I have raised a doubt about this, since it seems to assume that human moral psychology works in a manner that it quite possibly does not work. But this represents no great victory for the moral skeptic. Perhaps in cases of deep moral dispute, one is *not* required to bracket off so much that one cannot judge whether one's disputant is an epistemic peer, but if the disputant nevertheless disagrees with a great many of one's moral beliefs, then one may simply conclude that the disputant is not an epistemic peer. The difference is between "I can't tell whether you're an epistemic peer" and "You're not an epistemic peer." Either way, one lacks ground for taking the person to be an epistemic peer, and so conciliationism doesn't require a downgrade in the epistemic status of one's belief on the disputed matter.

But I think this continues to make the avoidance of moral skepticism a hostage to empirical fortune. It cannot simply be assumed that whenever someone disagrees with me on a certain weighty moral matter, this person must also disagree extensively on other matters so much so that I will deny that he or she is an epistemic peer. It's an empirical question to what extent human moral judgments can be compartmentalized. It might be responded that if Mary has reason to suspect that Fred's moral disagreement with her is due to some odd psychological compartmentalization on his part, then this disqualifies him from being considered an epistemic peer. But this seems contrary to the spirit of the epistemic modesty that conciliationism champions. Mary might well suspect that Fred's moral judgments are due to quirky aspects of human psychology rather than consistent inferences from more basic values (perhaps Fred's view on this one matter has been influenced by arbitrary factors), but epistemic modesty should leave her wondering whether her own heartfelt moral judgments are any better off. Mary may come to the conclusion that Fred is epistemically flawed in his moral judgments, but if she wonders whether she too is flawed in similar ways—that this is just how human moral psychology works—then this is no reason to deny that Fred is her peer.

It is also worth noting here that the kind of people who might disagree with a weighty and seemingly obvious moral truth do not need to be construed as moral monsters (Vavova mentions Caligula and Clarisse the "homicidal sociopath" [2014: 314])—they may instead be friendly and bespectacled metaethicists who happen to be error theorists about morality, or maintain skeptical views about certain elements of morality. In ordinary contexts, if someone disagrees with your claim that it is morally wrong to shoot strangers, then you'd take him to mean that it's morally permissible to shoot strangers, in which case you'd be appalled, think there was something wrong with him, and probably downgrade his epistemic credentials. But in the context of doing metaethics, this inference would be mistaken: an error theorist may deny that it is morally wrong to shoot strangers, but she should quickly append that nor is it morally permissible to shoot strangers—it's not morally *anything*—and she might add that she is adamantly and passionately opposed (on non-moral grounds) to shooting strangers. (Gilbert Harman [1975: 7] once claimed on metaethical grounds that we have no business saying that it was wrong of Hitler to have ordered the extermination of the Jews, though he quickly appended that we can say that Hitler was an evil man and that his actions brought about something that ought never to have happened.) Well, I suppose that one might still feel appalled at this, but it's a quite different sort of appallingness than the former kind, and it's considerably less obvious that espousing such a view should count immediately against one's epistemic credentials.

Conciliationism offers a more direct route from moral disagreement to justification skepticism than genealogical debunking arguments, but the route is far from trouble-free. One thing I've stressed is that how one

assesses another's epistemic credentials may differ in moral cases from non-moral cases, due to the psychological mechanisms lying behind moral judgment being substantially different from those that lie behind other kinds of judgment; and an interesting take-home message is that many of the relevant details remain unknown empirical territory.

6.   Conclusion

There are different kinds of moral skepticism, and various possible paths to each of these views. Many of those paths have nothing to do with the phenomenon of disagreement, and among these may be the strongest skeptical arguments. This essay has focused on several entwined arguments that do focus on moral disagreement, endeavoring to display their interrelations and difficulties.

Notes

1 Others have this preference also: see Brink (1984) and Loeb (1998). Charitably, one might assume that Mackie has in mind so-called *descriptive relativism*, but since I think that that is also a misleadingly labeled thesis, the charity is limited!

2 One way of plausibly denying P1 is to draw attention to the fact that realism is compatible with forms of relativism, and relativism does not predict convergence in moral opinion. On the other hand, moral relativism has some trouble accommodating the existence of disagreement: if, when I say "X is wrong" I mean *from point of view* φ, and when you say "It is not the case that X is wrong" you mean *from point of view* ψ, then what appears to be a disagreement turns out not to be one. So the possibility of relativistic realism complicates matters considerably, and for this reason I will bracket it off from this chapter. Besides, it doesn't appear to be a possibility that is on Mackie's radar; he seems to think of moral realism as necessarily an absolutist position. (Perhaps, like many, he confuses objectivism and absolutism.) This seems to me a flawed taxonomy (see Joyce 2015), but in this chapter I'll accept it for the sake of argument.

3 As Folke Tersman has pointed out (2015; Forthcoming). See also Enoch (2009: 22).

4 The parenthetical asides "(or objective moral facts)" indicate that there are two possible ways of reading Mackie here. We could read him as arguing against the existence of objective moral facts, and then presenting an argument that moral facts are essentially objectivist, thus arguing for a moral error theory. Or we could read him as *first* establishing that moral facts are essentially objectivist, and then just arguing against the existence of moral facts *simpliciter*. It's much of a muchness. I will go the latter route, assuming that what is under dispute is the existence of moral facts *simpliciter* (and thus I will henceforth do away with the parenthetical asides).

5 This amendment should dispel any worries one might have that (B) violates the aphorism "Absence of evidence is not evidence of absence." Absence of evidence *when it is reasonable to expect that presence would leave evidence* does count as evidence of absence.

6 Gilbert Harman (1977; 1986) is sometimes interpreted as claiming that moral facts do not play any explanatory role and therefore we should doubt (or reject) their existence. But he explicitly recognizes that the availability of a *naturalistic reduction* will save the day for moral facts. And he goes on to suggest particular

reductions that he finds plausible, ultimately coming to the conclusion that "there is empirical evidence that there are (relational) moral facts" (1977: 132). The moral facts that Harman accepts, though, are not objective in nature, so his view is not a realist one.

7 David Enoch writes: "A radically inaccessible realm of moral facts is, I think, a very small comfort for the realist. Such realism may, at most, serve as a last resort, but it is to be avoided if at all possible" (2009: 22). And Brink writes that if the realist were to rely too much on appealing to ties and incommensurability, it "would weaken his reply to the argument from disagreement" (1984: 116).

8 My focus on the identity relation is largely for brevity. In fact, other weaker nomological relations, such as supervenience or some kind of probabilistic casual relation, will suffice to undermine the debunking argument (see Brosnan 2011: 61). The difference is very important in other areas, but not, I think, to any of the arguments discussed in this chapter.

9 Tersman (Forthcoming) argues that consideration of moral disagreement may re-enter the debunking debate later: as a factor potentially counting against the moral naturalist's ability to defeat the debunking argument by providing a theory that plausibly connects moral facts to the non-moral facts accepted in the debunker's genealogy.

10 Others reject the conciliatory view in favor of the *steadfast view*: that it is acceptable to remain confident in one's beliefs in the face of disagreement from epistemic peers (see Kelly 2010; Sosa 2010; Lackey 2010). Here I am accepting a conciliatory view for the sake of argument.

11 Though Diego Machuca (2015: 27) has rightly argued that it is a mistake to read too much of the detail of the modern debate about the epistemic significance of disagreement into the views of the ancient skeptics.

References

Benjamin, D., Cesarini, D., Matthijs, J., Dawes, C., Koellinger, P., Magunsson, P., Chabris, C., Conley, D., Laibson, D., Johannesson, M., & Visscher, P. 2012. "The Genetic Architecture of Economic and Political Preferences," *Proceedings of the National Academy of Science* 109: 8026–8031.

Braddock, M. 2016. "Evolutionary Debunking: Can Moral Realists Explain the Reliability of Our Moral Judgments?" *Philosophical Psychology* 29: 844–857.

Brink, D. 1984. "Moral Realism and the Sceptical Arguments from Disagreement and Queerness," *Australasian Journal of Philosophy* 62: 111–125.

Brosnan, K. 2011. "Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge?" *Biology and Philosophy* 26: 51–64.

Choma, B., Hanoch, Y., Hodson, G., & Gummerum, M. 2014. "Risk Propensity among Liberals and Conservatives," *Social Psychological and Personality Science* 5: 713–721.

Christensen, D. 2007. "Epistemology of Disagreement: The Good News," *The Philosophical Review* 116: 187–217.

Das, R. 2016. "Evolutionary Debunking of Morality: Epistemological or Metaphysical?" *Philosophical Studies* 173: 417–435.

Elga, A. 2007. "Reflection and Disagreement," *Noûs* 41: 478–502.

Enoch, D. 2009. "How Is Moral Disagreement a Problem for Realism?" *The Journal of Ethics* 13: 15–50.

Ensminger, J., & Henrich, J. (eds.). 2014. *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective*. New York: Russell Sage Foundation.

Feldman, R. 2006. "Epistemological Puzzles about Disagreement." In S. Hetherington (ed.), *Epistemology Futures*, 216–236. Oxford: Oxford University Press.

Feldman, R., & Warfield, T. (eds.). 2010. *Disagreement*. Oxford: Oxford University Press.

Fraser, B. 2014. "Evolutionary Debunking Arguments and the Reliability of Moral Cognition," *Philosophical Studies* 168: 457–473.

Harman, G. 1975. "Moral Relativism Defended," *The Philosophical Review* 84: 3–22.

———. 1977. *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press.

———. 1986. "Moral Explanations of Natural Facts: Can Moral Claims Be Tested against Moral Reality?" In N. Gillespie (ed.), *Spindel Conference: Moral Realism (The Southern Journal of Philosophy Suppl. Vol. 24)*, 57–68.

Harman, G., & Thomson, J. 1996. *Moral Relativism and Moral Objectivity*. Oxford: Blackwell.

Hatemi, P., & McDermott, R. 2012. "The Genetics of Politics: Discovery, Challenges, and Progress," *Trends in Genetics* 28: 525–533.

Hibbing, J., Smith, K., & Alford, J. 2014. "Differences in Negativity Bias Underlie Variations in Political Ideology," *Behavioral and Brain Sciences* 37: 297–307.

Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.

———. 2006. *The Evolution of Morality*. Cambridge, MA: MIT Press.

———. 2015. "Moral Anti-Realism." In E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/archives/sum2015/entries/moral-anti-realism/.

———. 2016a. "Evolution and Moral Naturalism." In K. Clark (ed.), *Blackwell Companion to Naturalism*, 369–385. Oxford: Blackwell.

———. 2016b. "Evolution, Truth-Tracking, and Moral Skepticism." In R. Joyce (ed.), *Essays in Moral Skepticism*, 142–158. Oxford: Oxford University Press.

———. 2016c. "Reply: Confessions of a Modest Debunker." In U. Leibowitz & N. Sinclair (eds.), *Explanation in Mathematics and Ethics*, 124–145. Oxford: Oxford University Press.

Kahane, G. 2011. "Evolutionary Debunking Arguments," *Noûs* 45: 103–125.

Kelly, T. 2010. "Peer Disagreement and Higher-Order Evidence." In Feldman & Warfield 2010, 111–174.

Kornblith, H. 2010. "Belief in the Face of Controversy." In Feldman & Warfield 2010, 29–52.

Lackey, J. 2010. "What Should We Do When We Disagree?" *Oxford Studies in Epistemology* 3: 274–293.

Loeb, D. 1998. "Moral Realism and the Argument from Disagreement," *Philosophical Studies* 90: 281–303.

Machuca, D. 2015. "Agrippan Pyrrhonism and the Challenge of Disagreement," *Journal of Philosophical Research* 40: 23–39.

Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin.

Matheson, J. 2015. *The Epistemic Significance of Disagreement*. Hampshire: Palgrave.

Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Shafer-Landau, R. 1994. "Ethical Disagreement, Ethical Objectivism, and Moral Indeterminacy," *Philosophy and Phenomenological Research* 54: 331–344.

Slater, A., & Quinn, P. 2001. "Face Recognition in the Newborn Infant," *Infant and Child Development* 10: 21–24.

Sosa, E. 2010. "The Epistemology of Disagreement." In A. Haddock, A. Millar, & D. Pritchard (eds.), *Social Epistemology*, 278–297. Oxford: Oxford University Press.

Starkweather, K., & Hames, R. 2012. "A Survey of Non-Classical Polyandry," *Human Nature* 23: 149–172.

Sturgeon, N. 1985. "Moral Explanations." In D. Copp & D. Zimmerman (eds.), *Morality, Reason and Truth*, 49–78. Totowa, NJ: Rowman & Allanheld.

Tersman, F. 2015. "Review of R. Joyce and S. Kirchin (eds.), *A World without Values: Essays on John Mackie's Moral Error Theory*," *International Journal for the Study of Skepticism* 5: 333–337.

———. Forthcoming. "Debunking and Disagreement," *Noûs*. DOI:10.1111/nous.12135.

Vavova, K. 2014. "Moral Disagreement and Moral Skepticism," *Philosophical Perspectives* 28: 302–333.

Wielenberg, E. 2010. "On the Evolutionary Debunking of Morality," *Ethics* 120: 441–464.

8 Evolutionary Debunking, Realism and Anthropocentric Metasemantics

Mark van Roojen

1. Introduction

Evolutionary debunking arguments can be unnerving. If successful, they show an incompatibility between two commitments that are pretty deep in my world view. I'm rather sure that I know that certain sorts of actions are wrong. And I'm sure that we are as we are as a result of certain natural evolutionary processes, at least roughly of the sort Darwin hypothesized. So it should be a comfort to agree with prominent debunkers such as Sharon Street who claim that such arguments favor anti-realist or constructivist theories over completely mind-independent accounts of morality. It already strikes me as implausible that the nature of morality would have nothing at all to do with the nature of human agents. If Street is right, my minimally constructivist sympathies would insulate my views from such debunking. I could then just heave a sigh of relief and be grateful that my own commitments are not in tension.

As it stands, however, I think these debunkers are wrong to see their arguments as raising a special problem for realism. In a nutshell, I think they conflate mind-dependent content-determination relations with mind-dependent content. Or to come at the point from a different direction, insofar as reference is broadly an epistemic relation, evolutionary debunking arguments would cause trouble for mind-independent theories of reference and content-*determination*, since there would be no guarantee that reference would track epistemic access. But a firmly realist theory of content is consistent with a mind-dependent theory of reference and content-determination.[1] To use a toy example, most facts about rocks are mind-independent, but we are only in a position to refer to and talk about rocks because of the faculties that enable us to know about rocks. This should seem platitudinous, but it turns out to have important implications for evolutionary debunking arguments. These are (1) that fully realist externalist naturalist views are in no real danger from these arguments; and (2) that a certain sort of non-naturalist view might also be able to use a parallel strategy to resist debunking; but (3) this will be somewhat harder to do given one of the thoughts that often motivates non-naturalism. I will argue for these claims

by presenting a model for a realist theory of content that seems untouched by evolutionary debunking. Finally, I express a worry that internalist views of the sort I myself favor may have problems similar to those I raise for non-naturalism, whether they are naturalist or not. But first some setup.

2.   Debunking Deployed Against Realism

I'll begin with a sketch of a relatively generic[2] evolutionary debunking argument, presenting it in an order I find intuitive:

1.   The range of moral judgments we actually make is significantly dependent on our affective responses and, in turn, on our capabilities and dispositions to have such responses. (Affect effects judgments.)
2.   Our affective capabilities and dispositions are as they are as a result of a process of random variation, culling and inheritance. The culling process tended to allow dispositions to be inherited only if they were relatively advantageous for survival in the environment where they occurred. Relatively disadvantageous variations tended to go to extinction through the same process. (Affective capacities evolved.)
3.   Therefore, if our moral judgments come anywhere close to tracking the truth, it must be because this process of variation, culling and inheritance shaped our affective dispositions so as to allow our judgments to do so. (Truth tracking must be selected for.)
4.   If the moral truths are independent of our psychological capacities and dispositions (or of adaptiveness in our environment), it is highly unlikely that such a random process would have caused us to have capacities and dispositions that allow our judgments to come relatively close to tracking those truths. (No selection to track affect-independent truths.)
5.   Therefore, we should conclude either that we are likely not coming relatively close to tracking truths with our moral judgments or that the moral truths we track are mind-dependent in some significant way.

The argument has three premises: steps 1, 2 and 4. The first two premises, properly understood, are true. The third step, properly understood, follows from the first two. Since I dissent from the conclusion, I must either dissent from the remaining premise, step 4, or argue that the conclusion does not follow from it. I think there is a reading of the fourth premise on which it is true, but that the conclusion does not follow from it. There is also a reading from which the conclusion does follow, but that reading of step 4 is not in fact true. The former deploys an antecedent probabilistic notion of likelihood, whereas the latter reading is about current epistemic probability. I'll leave that cryptic claim alone for now, coming back to explain it after I sketch my response on behalf of realism later in the essay.

   I should probably explain how I understand the two premises I accept full stop and why they're true so understood. The first postulates a causal

influence running from our affective dispositions to the range of moral and other normative judgments we actually make. It does not say this is the only influence, or that these dispositions by themselves determine us to make the judgments we do. But it does imply a certain counterfactual dependence—if we lacked some of the affective capacities and dispositions we have, we would not make a wide and important range of the judgments we now make and believe. For example, without some empathetic understanding of other people and other creatures, we would be unlikely to endorse altruism to the extent we do. Without a disposition to value reciprocity, we would be unlikely to judge friendship an important value and to find fairness important. And so on. So understood, the first premise should be uncontroversial. The second premise recognizes that these capacities are to a large extent contingent and that we have them as the result of evolution, which at one level involves random variation. To the extent that there is selection among traits, it is not intentional selection. Rather, it is a process where the traits which are more disadvantageous than alternative traits present in a species lead creatures with those traits not to do as well as those without them, and hence to leave fewer progeny with the disadvantageous trait. The traits that are more advantageous thus have a greater tendency to persist. The upshot is not that traits which persist are optimal, only that they were more adaptive than the available alternative thrown up by random variation.

Taken together, these claims imply that any affective dispositions that enable us to track the truth about morality must be the result of this process of variation and filtering. That's the third step in the argument, which follows from the first two. Sharon Street, most prominently, has defended the fourth step (and third premise) of the argument. She argues that realists about value (as opposed to anti-realists) cannot explain the congruence between "our evaluative attitudes, on the one hand, and the independent evaluative truths that realism posits, on the other" (2006: 109). Bedke (2009; 2014) similarly argues that any congruence would be a cosmic coincidence. We'll need to look at Street's more detailed argument, but first we need to get clearer about the realism/anti-realism divide. Everybody thinks that some moral truths are attitude-dependent. One reason one shouldn't insult people is that it makes them feel bad. Street suggests we can get a clean divide between realists and anti-realists if we look at whether a theory posits any judgments whose truth or falsity is determined independently of the whole set of evaluative judgments we do make or might make upon reflection as well as any other evaluative attitudes we hold or would hold upon reflection. Realists will think that there are moral truths whose truths do not depend on these attitudes, whereas anti-realists in her sense will think all moral truths have this sort of dependence (Street 2006: 111). This way of dividing realism from anti-realism is supposed to come to the same thing as dividing "stance-independent" realist truths from "stance-dependent" irrealist claims (Street 2006: 111; Milo 1995: 182; Shafer-Landau 2003: 15). So understood, it looks like certain sorts of naturalists—those who don't

reduce moral truths to facts about human psychology—and non-naturalists will count as realists.

Irrealists, Street suggests, will have an easy time explaining the congruence of moral truth with the judgments that our affective capacities and dispositions allow us to make. Rather than explaining why our capacities evolved so as to allow us to track an independent set of truths, they can reverse the order of explanation. They can argue that insofar as our attitudes enter constitutively into the nature of the truths in question, it is no surprise that those truths will co-vary with the relevant attitudes. So, for example, if the moral truths just are the judgments we would stably make under conditions of full information and imagination, including empathy, it will be no surprise that when we have such information and are sufficiently imaginative and empathetic, we'll have some idea of what is in fact right and valuable.[3]

Clearly realists won't be able to give this kind of explanation. There will be no non-trivial condition under which our judgments about value must be right. So they cannot define rightness in terms of those judgments to secure the requisite correlations. But that is not to say that they might not have other explanations available. In fact, as I will soon argue, there is a perfectly good explanation of how evolution could put us in a position to know about realism-friendly mind-independent normative properties. If evolution in fact does that, the reading of premise 4 that supports the conclusion in my reconstruction of the anti-realist argument will be false.

But it isn't obviously false, so we should look at some arguments to the effect that premise 4—the claim that realists cannot explain the congruence of our inherited evaluative dispositions with moral truth—must be true. In her original presentation of her argument, Street suggests that either natural selection of evaluative dispositions tracked the moral truth, or it didn't. If it didn't, then while occasional moral judgments might by chance turn out to be true, there would be no general tendency for our moral beliefs to be close to true. Even when corrected by deliberation and reflection, evaluative dispositions which are only ever aligned with the moral truth by chance will be no likelier to help us get things right as to get things wrong. That's because these processes will have at least as many misleading as veridical judgments to work from (Street 2006: 124).

On the other hand, if there is some relation between the affective tendencies favored by evolution and the independent normative truths, and if that relationship is positive rather than negative, there might be some hope of vindicating the judgments made available to us by evolution. Street suggests that the required relation would have to be one in which we evolved the abilities to make true judgments because true judgments contribute to reproductive success. She thinks that the true judgments would have to be conducive to reproductive success "because they are *true*, and it proved advantageous to grasp evaluative truths" (2006: 128). The problem for realists on this horn of the dilemma is that there is a better explanation of how we came to

make these judgments that doesn't posit their truth. This explanation suggests that creatures who believed (some of) the judgments that our evaluative dispositions enable us to make tended to act in accordance with those judgments. And acting in accordance with those judgments tended to favor reproductive success (2006: 129). Absent some showing that this tendency is itself to be explained by the moral truths, we have a more parsimonious explanation that does just as well explaining these dispositions.

There is at least one widely noted problem with this argument: it isn't obvious that realists need to say that the ability to make true moral judgments was adaptive because those judgments are true. There could be a non-accidental connection between the ability to make judgments that are as a matter of fact true and reproductive success in an environment. This could favor creatures who are able to make such true judgments, but not because they are true. So-called "pre-established harmony" or "third factor" (Enoch 2010) accounts posit just such an explanation. Survival is good as is reproductive success, and not accidentally so (Enoch 2010: 168ff.). Creatures with certain cognitive capacities necessary to being able to make moral judgments necessarily have rights (Wielenberg 2010). Well-being is good (Brosnan 2011). Pleasure is intrinsically good (Skarsaune 2011). And it is not an accident that evolution would select for the belief that survival and reproductive success is good, or that creatures like us are protected by a moral barrier, or that pleasure is good, or. . . . Defenders against evolutionary debunking can claim that it is not at all implausible that we evolved to have these very beliefs, so that given the necessary truth of these moral claims, it is not an accident that we believe the truth. I think there is something to these replies,[4] but there remain some worries. One has to do with parts of morality that don't seem explicable in this way. For instance, it isn't obvious that being right about survival, reproductive success and rights is enough to generate any particular view about our duties to animals. Another worry has to do with the thought that knowledge requires believing something because it is true,[5] and these explanations don't seem to secure that. Still, I'm not dismissing this strategy of response, partly because I'm not sure my own actual views can take advantage of the response strategy I am about to offer the thoroughgoing realist.

3. A Strategy of Realist Response

The thought that realism cannot explain how we come to accurately represent moral reality given the stance-independence of moral properties seems very compelling until one distinguishes the nature of the properties themselves from the semantic theory, which explains how we come to be able to talk about these stance-independent properties. The properties we refer to can be perfectly objective stance-independent properties while our ability to refer to these properties can depend on all sorts of contingencies about our natures. I will argue for this claim by giving a model from another domain that has

both of the features I claim a realist theory of morality could have. With that example in place, I will show how it is compatible with thinking it antecedently unlikely that we came to have knowledge of these very properties, while still thinking that it is now very plausible to think that we have such access.

My model comes from a theory of colors developed by David Hilbert (1987). He calls the view "Anthropocentric Realism."[6] On this view, colors are objective observer-independent properties of objects, roughly their surface spectral reflectances. Color "is the disposition objects have to reflect varying percentages of incident light" (1987: 119). That an object has such a disposition is completely independent of observers, whether actual or possible. It likely depends on the microstructure of the surface and the nature of light, but not on anything subjective. Therein lies the view's realism. The anthropocentrism comes in to explain which such dispositions we refer to with our color terms. A first level of anthropocentrism is present insofar as different languages divide color space differently from one another (1987: 130). We can discriminate many more colors than we have words for. Our color terms pick out classes of similar reflectance properties that look similar to us, but where one class begins and another one ends is somewhat a matter of choice. Once the boundaries are decided, however, the classes themselves are perfectly objective. As Hilbert notes, we could design a device to sort token color samples into the classes. The groupings are to some extent up to us, but that doesn't mean that our terms don't pick out real properties that would be instantiated or not whether we had existed or not.

A second level of anthropocentrism comes into the picture because we have three kinds of color receptors, each sensitive to light within a certain range. Colors are computed on the basis of the relative intensities of light in the ranges to which these receptors are sensitive. This means that two surfaces with different reflectance profiles will look the same to us across a range of normal lighting circumstances, so long as the peak reflectances are indiscriminable by each sensor. This is the phenomenon of metamerism: "Metamers do not differ just slightly in their reflectance profiles. An object with an essentially flat reflectance curve can appear to have the same color as one with large peaks in reflectance" (1987: 103). Our color terms, then, will group reflectance patterns together as cardinal red (say) just when they affect our receptors in the same way. And in one sense very different reflectance properties can do that. Still, Hilbert suggests, we can offer a unifying account of the relevant properties of a sort:

> [H]uman color vision is . . . a passive system with three types of broad-band sensors. As we have seen, objects that reflect the same amounts of light in each of the wave-bands will appear to have the same color. We can express the common property that the reflectances of such objects will share by summing the reflectances of the objects over each of the three ranges. We will obtain what is called a triple of integrated reflectances.
>
> (1987: 111)

Once again, these are real properties that would be there whether we existed or not. But our talking about them is a function of what we are like. Why does our color talk refer to triples of reflectances at just these bands? Because our visual system has sensors that are sensitive at just these. Other creatures could and probably do have visual systems like ours but sensitive at different bands. There might be creatures with more or fewer broadband receptors. If there were such creatures and they could talk about their visual experiences, they would, according to this theory, refer to different reflectance properties than we do. But for all that, the reflectance properties that would be the referent of their "color" terms would be perfectly objective. We are presently surrounded by them. We just don't discriminate them as those creatures would and hence are not (usually) in a position to talk about them.

My use of the model does not really depend on its correctness as an account of our color terms, though I find it pretty plausible. It presents a model for stance-independent realism that makes it less mysterious how we could be in a position to get the facts right without introducing any stance-dependence into the nature of the properties picked out. Hilbert's color realism identifies those properties with the perfectly objective physical properties to which we are sensitive. But our being sensitive to them is not part of the nature of the properties themselves, nor are those properties in any way dependent on us. However, the story of how we come to be in a position to talk about these properties—how our color terms come to pick out just these properties—does depend on our nature. This means that the semantic values of the color terms just are mind-independent properties. But the metasemantic explanation of how our terms come to have these referents makes essential reference to our ability to discriminate these properties, and that means that the full story will get into the contingent details of our visual system. On a Millian version of the metasemantic story, we would begin with some population of people who are able to distinguish the relevant reflectance properties by how they look to them and who coin terms to pick out certain classes of them. These are then passed on to a wider community of users, who use them to refer to the same properties as the original bunch did. On this story, the actual ability to talk about these properties depends on the nature of the observers, but the properties themselves don't. And, since it is the properties that are picked out by the color terms, there is no anthropomorphism in the contents of sentences using color vocabulary. A yellow rose would be and remain yellow, even if our visual capacities and dispositions changed. We would rather lose our ability to discriminate yellow from other colors.

There is a lot of evolutionary contingency in the story. Probably the exact bands of light to which our receptors are sensitive could easily have been otherwise, and we might have had more or fewer receptors (though not many more or many fewer).[7] When our sightless ancestors started the process of evolving into us, it was a long shot that we would see the colors we do. It is improbable that the exact shades we do in fact see are such that we

were selected to see these and not some other set within the range that is illuminated on earth. So there is no tracking explanation for our seeing just these colors rather than some other colors.[8] Yet that should not make us skeptical of our ability to see the colors we see. It is no miracle that we are sensitive to surface spectral reflectance, and being sensitive to it will require some color perception system or other. Any one of these would have been a long shot. So it should be no surprise that the one we wound up with was one of these long shots.

4.   A Similar Model for Value Realism

Realists about values and morality of the sort Street targets think our moral terms function to pick out perfectly objective stance-independent properties. I take it that most of the Cornell realists qualify. I will use Richard Boyd's (1988) account as a basis to sketch a view of value parallel to Hilbert's views about color. Boyd thinks that moral goodness is a homeostatic cluster property—that is, a property composed of a number of properties which, when co-instantiated, tend to keep each other in existence. The model for these properties is health in animals, including human beings. Health has a number of components and a healthy creature is such that the presence of one component tends to reinforce the presence of the others. These properties are perfectly stance-independent. They would be instantiated where they are, whether or not we had ever come to think and talk about them.[9] Reference to a value property is secured by an epistemically virtuous feedback loop, leading from the property to what we think and say about it and to how we interact with it, which in turn provides us information we can use to modify our views of the property. Boyd emphasizes the epistemic nature of the relation:

> *Roughly*, and for nondegenerate cases, a term *t* refers to a kind (property, relation, etc.) *k* just in case there exist causal mechanisms whose tendency is to bring it about, over time, that what is predicated of the term *t* will be approximately true of *k* (excuse the blurring of the use-mention distinction). Such mechanisms will typically include the existence of procedures which are approximately accurate for recognizing members or instances of *k* (at least for easy cases) and which relevantly govern the use of *t*, the social transmission of certain relevantly approximately true beliefs regarding *k*, formulated as claims about *t* (again excuse the slight to the use-mention distinction), a pattern of deference to experts on *k* with respect to the use of *t*, etc. . . . When relations of this sort obtain, we may think of the properties of *k* as regulating the use of *t* (via such causal relations), and we may think of what is said using *t* providing us with socially coordinated *epistemic access* to *k*: *t* refers to *k* (in nondegenerate cases) just in case the socially coordinated use of *t* provides significant epistemic access to *k*, and not to other kinds (properties, etc.).
>
> (Boyd 1988: 195)

Whether this is a good account of reference in general or moral reference in particular is not presently my concern. What I want to note is that on this account, we must have epistemic access to the properties that are the referents of our terms.

This means that according to this kind of realist theory, the referent of our terms must be something we have epistemic access to. And that means that the reference-determination story will bring in a kind of stance-dependence, not of the property which is the referent, but of our ability to refer to it. I think this is a general feature of theories of reference, but for now I'll stick with my example. Suppose now that the realist takes the first horn of Street's dilemma, and suggests that evolutionary processes did not select for correct moral judgments. She suggests:

> [W]e are left with the implausible skeptical conclusion that our evaluative judgements are in all likelihood mostly off track, for our system of evaluative judgements is revealed to be utterly saturated and contaminated with illegitimate influence. We should have been evolving towards affirming the independent evaluative truths posited by the realist, but instead it turns out that we have been evolving towards affirming whatever evaluative content tends to promote reproductive success. We have thus been guided by the wrong sort of influence from the outset of our evaluative history, and so, more likely than not, most of our evaluative judgements have nothing to do with the truth. Of course it is possible that as a matter of sheer chance, some large portion of our evaluative judgements ended up true, . . . but this would require a fluke of luck.
>
> (2006: 122)

This is not what we should say if we accept Boyd's reference-determination story. Were we in a scenario where most of our beliefs about the relevant evaluative properties were false, and if we lacked a feedback mechanism that allowed us to correct these false beliefs over time, we would not be referring to these properties with our evaluative terms. We would either be referring to a different set of properties—properties which do stand in the requisite epistemic feedback relation to our use of those terms—or we would be referring to no properties at all.

Hilbert's color theory provides a parallel. It is unlikely that we were selected for our ability to see fuchsia. It is also unlikely that at the beginning of sighted existence, our current visual apparatus was anything more than one of many ways that creatures might implement a light-sensitive perceptual system. More probably, any one of a range of possible implementations was a real possibility,[10] and the accidents of our evolutionary history explain why we have the one we do. But we have the one we have and we can discriminate fuchsia. Suppose things went differently, so that our actual visual system did not group all and only the reflectance profiles we see as fuchsia together. This would be a world in which some things were still

fuchsia since there would still be the reflectance profiles picked out by our present term 'fuchsia'. But we would not see colors in this profile as similar to one another. And we would not have a word for fuchsia and would not talk or think about fuchsia. If we used a term phonetically and orthographically like 'fuchsia', it would pick out something other than fuchsia.

This does not mean that we, as we are now, have benefited from some tremendous fluke of luck to be able to get things right about fuchsia, about which we would otherwise have been wrong. True, the antecedent chances of our developing just this visual system were very small. The discriminatory abilities which underwrite the epistemology of color also secure our ability to refer to the colors we do. The contingent fact that we evolved to have these abilities is responsible for both our ability to talk about fuchsia, and our knowing some things about it. That we can do both rather than only one is not a fluke. What is in some sense a fluke is that we perceive precisely fuchsia, and the other colors we see. But whatever colors we would have been in a position to discriminate, it was antecedently a long shot that we would evolve to be able to know about them.

Boyd should say the parallel thing when it comes to the dispositions that underwrite our evaluative judgments. Yes, it was not necessary or perhaps even likely that we wound up in a position to grasp the evaluative truths we recognize. We might have been different. But then we would not have been talking about the same evaluative properties we presently do talk about.[11] We might have been talking about some other properties, perhaps similar enough to our present properties that we should think of them as evaluative as well.

At this point, it would be natural to respond that this is a form of relativism and that relativism is a form of anti-realism, not realism. Street in fact says something like this:

> In order to count as genuinely realist, then, a version of value naturalism must take the view that *which* natural facts evaluative facts are identical with is independent of our evaluative attitudes. For ease of expression, let us put the point this way: in order to count as realist, a version of value naturalism must take the view that facts about *natural-normative identities* . . . are independent of our evaluative attitudes.
>
> (2006: 137)

I think that this is a mistake, though an easy one to make. The postulated theory does not say that the truths would change depending on which attitudinal dispositions we had evolved to have. It only says that our present sentences would not have picked out the same propositions under alternative scenarios. Most of the same propositions would remain true (though not the ones about our evolutionary history). We would just be ignorant of them. The point is perfectly general. If we never had epistemic access to the things we think and talk about, we would not have been able to talk about them.

This holds for evaluative things, color, the Big Bang, rocks and trees. This isn't relativism, it is just an ordinary fact about reference-determination that falls out of Boyd's metasemantic theory.[12]

Relatedly, when this kind of naturalist identifies evaluative properties with natural properties and thus evaluative facts with natural facts, these identities are not supposed to be creatures of our thought or language. Everything is necessarily self-identical. It is possible that in some sense we might have picked out different properties with our present evaluative terms, and perhaps these properties would have been similar enough to those we presently use our terms for that we should call those evaluative as well. Nonetheless, we would be talking about different properties and facts than we presently do—at least if something like this theory of reference-determination is correct.[13]

At this point, it might be very tempting to pull out Horgan and Timmons's (1992) Moral Twin-Earth (MTE) argument to argue that we would take our counterfactual evolutionary counterparts with different evaluative dispositions to disagree with us when they make judgments consistent with their less empathetic moral dispositions.[14] The thought experiment is, after all, aimed at Boyd's semantic and metasemantic theory, since it predicts that there will be no real disagreement when our judgments are appropriately causally and epistemically related to different referents. There are several things to say here. The first is that semantic externalists in broadly the same camp as Boyd wish to deny the probative value of those thought experiments (Schroeter & Schroeter 2014; Dowell 2016). Secondly, those thought experiments are themselves viewed by their proponents as bringing in the kind of stance-dependence that Street thinks her argument from evolution shows we need. Horgan and Timmons use the argument precisely to argue for a species of non-cognitivism according to which all moral judgments must be made from an affectively engaged stance. If they're right, it isn't obvious that the evolutionary argument is now adding anything.

Furthermore, it is a delicate matter exactly what the Moral Twin-Earth argument shows. I think it does support a version of moderate morals/motives internalism.[15] But it does not, I think, show that semantic externalism of roughly the sort Boyd proposes is incorrect.[16] If I'm right, realists who think the MTE argument is no reason to abandon their semantic theories have no other reason not to avail themselves of this response exploiting the epistemic character of reference-determination. And if they do so avail themselves, they will have a metasemantic theory which makes it the case that moral terms refer only to those things to which they stand in an appropriate epistemic relation. On the other hand, if the MTE scenarios really show that no externalist metasemantics is possible, and if only such metasemantic views deliver the upshot that reference is a kind of epistemic relation, my overall point would be interesting but ultimately of no use to a realist.

5. Connecting the Response to the Original Debunking Argument

I said earlier that the fourth step of the generic debunking argument with which I began had a true reading that did not support the conclusion and a false reading which did. I can now explain. That premise was:

4. If the moral truths are independent of our psychological capacities and dispositions (or of adaptiveness in our environment), it is highly unlikely that such a random process would have caused us to have capacities and dispositions that allow our judgments to come relatively close to tracking those truths. (No selection to track affect-independent truths.)

It is true that we might easily not have been in a position to track truths involving these particular properties. That claim is true. But from this, the conclusion stated in step 5 does not follow:

5. Therefore, we should conclude either that we are likely not coming relatively close to tracking truths with our moral judgments or that the moral truths we track are mind-dependent in some significant way.

It does not follow because, had we not tracked these particular truths, we would have been talking about something else about which we might still be in a position to know quite a lot. Given the way we evolved, the phenomena we are talking about are the phenomena our evolutionary history put us in a position also to know about; we can refer only to those things we have in our epistemic grip. So we are at least on track to know some things about values and morals. But the fact that *reference-determination* is mind-dependent does not show that the properties picked out in this mind-dependent way are mind-dependent.[17] So both options offered by step 5 can be false, consistent with the true reading of step 4.

6. Can Non-Naturalist Realists Make Use of the Reference-Determination Strategy?

I've offered a response to evolutionary debunking arguments on behalf of naturalist realists. Can non-naturalist realists adopt a similar response? I find it really hard to say, partly because the commitments of non-naturalism are pretty slippery, and because it is hard to say what all and only the natural properties have in common and thus what commitments separate naturalists from non-naturalists. Given the existence of non-reductive naturalists such as Sturgeon (2006), we might wonder what naturalists and non-naturalists are arguing about. Non-reductive naturalists and non-naturalists agree that moral properties are further properties over and above the natural properties on which they supervene. And they can agree that the relations of the

higher-level normative properties of actions or objects to the lower-level properties of those actions or objects are very much like the relations of biological or other special science properties of objects and the properties of lower-level sciences like physics. Yet, so-called "robust" realists like Enoch (2011) seem to differ from naturalists mostly by insisting that the relevant properties are "just too different" from ordinary natural properties to be of the same general type. We might then see their disagreement as being over whether normative properties belong in a natural kind of natural kinds— the natural. If the natural properties themselves form a natural kind, there is some real similarity that they all share. Naturalists think that normative properties too share this real similarity and hence belong in this higher-level natural kind of kinds. Non-naturalists disagree because they think that normative properties are "just too different."

What then might these differences be? Some non-naturalists insist that such properties are simple, whereas others (McDowell 1985; Scanlon 2014) suggest that normative properties have a kind of relational structure. But simplicity doesn't seem to be the right kind of difference. Some perfectly natural properties are simple, and in any case non-naturalist adherence to that commitment seems to be a historical accident stemming from Moore's (1903) use of simplicity to explain unanalyzability along with his conflation of the nature of our thought about the properties with the nature of the properties themselves. Many non-naturalists deny the causal efficacy of non-natural properties, and Bedke (2009, 2014) takes this as constitutive of non-naturalism in his debunking argument.

Denying causality to non-natural properties would seem to rule out a flat-footed application of Boyd's causal regulation semantics. But it isn't obvious to me that non-naturalists should abandon direct reference accounts that depend on some sort of epistemic feedback from the property. First off, if the worry is that normative properties are Platonic universals and that such universals can't cause anything (Enoch 2010), Platonism is a very general view and if it is plausible, it would apply to all properties, natural and non-natural alike.[18] This version of the objection would show that no properties can cause anything, not something any view should take on board. In any case, all Boyd's theory really needs is that events in which the relevant properties are instantiated cause other events so as to give us evidence about the nature of the properties instantiated. This may suggest a different objection: perhaps the thought is that normative properties are epiphenomenal, supervening on underlying properties which do the real causal work. But the natural properties favored identified with moral properties by naturalist moral realists also supervene in just this way. Homeostatic cluster properties supervene on and are constituted by the properties physics typically works with. The causal impotence of higher-level phenomena would undermine both naturalist and non-naturalist causation and hence causally mediated reference-determination. The views seem to be on a par. It would be more plausible in any case to follow common sense and let higher-level

phenomena inherit the causal powers of the things that compose them. Baseballs can break windows because their molecules can. Mental states can cause action because the physical states that realize them can, and so on. But if non-naturalists take that view, they should give up on causal impotence, as Russ Shafer-Landau (2012) seems willing to countenance.

So it would seem non-naturalists too could adopt a strategy parallel to the one offered to naturalist realists. They too should say that had we evolved to make sufficiently different moral judgments than we actually do, we would have epistemic access to different properties and hence be talking about those different properties. This requires the possibility of different non-natural properties as possible referents for the judgments made in those counterfactual circumstances. But without saying more about what makes these properties non-natural, I don't see any in-principle reason there could not be. If different higher-level natural properties can supervene on lower-level physical properties, and if therefore instances in which they are instantiated give us causal-epistemic feedback, we would have a picture parallel to the naturalists' view, substituting non-natural for supernatural properties.

But this doesn't seem to me to be much in the spirit of many if not most non-naturalist views. I think there's another motivation for the "just too different" thought. And this motivation would not sit comfortably with allowing that had we evolved differently, we would have referred to different properties with our moral and evaluative terms. My hypothesis is that many non-naturalists think that the connection such properties have with reasons—their normativity—makes moral and evaluative properties special, and that none of the natural properties have that same connection. One way to cash out the thought is as a kind of morals/reasons existence internalism[19]—moral and evaluative properties would have to be necessarily reason-giving, at least to us. If we couple that idea with the suggestion that a sufficiently different evolutionary endowment would have caused our evaluative and moral terms to track different properties than they actually do, we are faced with an obvious question. Would those different properties have been genuinely normative? In other words, would they be reason-giving? If the answer is no, we do, I think, wind up back in the grip of a worry about coincidence. How are we so lucky as to be the ones who track the only genuinely normative properties around? (And similarly, how are we so lucky as to be tracking the only non-natural properties around?) If we take the other option and say that those properties too are genuinely normative and reason-giving, we once more see an obvious question. Are they reason-giving to one and all, or just to people with these alternative evaluative dispositions? If the answer is "to one and all," we will likely have conflicting reasons stemming from the two sets of normative properties with non-tracking extension. That doesn't seem like an attractive view. Better to say that each property is normative for those with the disposition that allowed their judgments to track those properties. This looks a lot like stance-dependent normativity. And now the non-naturalist view winds up

not being realist in Street's sense.[20] And yet it looks like the cosmic coincidence worry has some force for those who hold it.

Cornell realists don't have the same issue because they are already quite happy to give up on morals/reasons internalism. Their strategy has typically been to deny that moral properties are necessarily reason-giving (Brink 1989). They have instead attributed a weaker status to moral properties, that of being interesting to us much of the time. Their strategy has generally been to explain why internalism was plausible without postulating a necessary connection between reasons and moral properties. If humans would often enough be motivated by goals that were attractive because of their moral attributes, externalists could agree that these properties are reliably connected with our reasons and motives without taking that connection to be necessary (Railton 1993). Were we to have evolved differently, so that our evaluative dispositions caused us to fix on other properties as most relevant for practical purposes, we would, as the debunkers argue, have made different moral judgments. But we would also have valued different things so that those things would in that case also have the reliable but not necessary connection with action that the morally valuable things now actually do. Or so the Cornell realists should argue.

7. Where Are We and How Did We Get Here?

We have seen that a property can be completely objective and stance-independent in the sense that it would exist whether we were in a position to know about it or not, even while our ability to think and talk about it does require us to be in a position to know about it. David Hilbert's theory of color presents a model for such properties and reference to them. And we have explored how an externalist metasemantics for moral terms like Boyd's causal regulation theory can exploit this possibility to explain how we in fact manage to refer to such properties despite it being unlikely, antecedent to our actually having evolved to refer to these particular properties, that we would wind up referring to precisely those properties we do. Furthermore, since our being able to think and talk about them depends on our being in the relevant favorable epistemic position with respect to these properties, such a view eludes the argument from the unlikelihood of our having come to have just the dispositions that put us in this favorable position to the unlikelihood of our now getting things right. Under the relevant skeptical scenarios, we would have been talking about different properties about which we would then be getting things right. So Street is wrong to think that stance-independent realist views are especially subject to her debunking argument.

Nothing in the Boyd metasemantics places all that many conditions on the nature of the properties referred to, beyond their being able to generate epistemic feedback through some causal powers or other. This opened up the possibility that non-naturalists willing to countenance causally

efficacious non-natural properties could adopt a similar metasemantics and answer evolutionary debunking arguments in parallel ways. In fact, I claim, this possibility is real. But it is in tension with one motivation for non-naturalism. This motivation is that normative properties are special—too special as the non-naturalists think of it to be natural. There are different ways to cash this out, but one way might be to say that they are the only properties that could be genuinely reason-giving.[21] But if we say that, then had we tracked any other properties with our normative terms, they would not have been genuinely normative. Only some epistemic feedback loops would put us in touch with genuinely normative properties. And now this non-naturalist is again saddled with the cosmic coincidence worry that the Cornell Realist was able to parry, for we have no explanation of how we came to be so lucky as to track the only genuinely normative properties.

That's where the argument has taken us so far. But we should notice that non-naturalness itself has played no real role in the argument. What did the work was the idea that the properties we track are genuinely normative and reason-giving, whereas the properties we would have tracked had our randomly selected propensities been different would not have been genuinely normative. A non-reductive naturalist might be tempted by that thought as easily as a non-naturalist. As might someone like myself, who is currently agnostic between these non-reductive options. I admit to being tempted, but this line of thought has me thinking twice.[22]

Notes

1 A point very well made in a more general way in Schroeter & Schroeter (ms).
2 In this generic form, it is in the ballpark with arguments by Street (2006; 2008; 2011) and Bedke (2009; 2014), and perhaps Joyce (2006; 2016). My discussion below will focus more on Street's views since her targets are all response-independent views, whereas Bedke just aims at non-naturalist views. I'll say something about how such arguments might be resisted by non-naturalists at the end of the chapter.
3 Tropman (2014) makes the case that the advantage to irrealism/constructivism is illusory by arguing that such views have their own problems.
4 I used selection for limited altruism to make a similar point in a 2006 American Philosophical Association symposium on what became Street (2011).
5 Whatever that elusive "because" comes to. Discussions of sensitivity and safety (for example Clarke-Doane 2011; 2012) strike me as attempts to partly elucidate the notion, but I suspect there is more to it.
6 This is far from the first metaethics paper to exploit an analogy with color. Phillip Pettit's Tanner Lectures (2015: 215), for example, deploys it to point out something related to my point—that anthropocentric properties can be both natural and real.
7 Most non-primate mammals have two; reptiles, amphibians and most birds have four; and pigeons have five. https://en.wikipedia.org/wiki/Color_vision#In_other_animal_species.
8 There's not really an "adaptive link" explanation for that either, if I'm right that some similar but different set of receptors would have served us just as well where survival is concerned.

9 Boyd suggests that with all scientific properties, it is a matter of luck whether we come to be in the appropriate relation to refer to them.

10 What range? My not very educated guess is that the most probable systems would have had more than two but not more than five broadband sensors and that the sensors would all have been sensitive to some range of the light actually present in our environment. But the specific boundaries of these sensors would be somewhat up for grabs.

11 Shafer-Landau (2012: 11) makes a suggestion that is similar but different: "[C]ertain moral propositions are reference-fixing, such that denial of (enough of) these propositions shows that an agent is no longer talking about morality at all." His thought is that people who make different judgments about certain moral paradigms won't exercise "semantic competence" and hence won't be talking about morality. This makes getting these things right a "conceptual truth" about the senses of the relevant moral terms. It is compatible with Boyd-style metasemantics that people believe all sorts of crazy things about the referents of their terms, just so they have an epistemic pathway to learning the truth. FitzPatrick (2015) and Vavova (2014) make similar suggestions to Shafer-Landau's.

12 See Schroeter & Schroeter (ms: n. 21) for the same point pressed against Street.

13 I think similar results would fall out of most theories of reference-determination, but I won't argue for that here.

14 This is in fact what Street does when she considers the possibility of "rigidified" naturalism.

15 Moderate morals/motives internalism holds that it is necessary that in some range of normal cases, moral judgments motivate those who make them. See Dreier (1990) and van Roojen (2010; 2017) for the connection with the MTE.

16 Boyd's theory can be modified to take the MTE argument into account while preserving the referential nature of the semantics. See van Roojen (2006). It would be too difficult to work what I say there into this chapter, but it may be important to note that the problem for Boyd in the MTE story does not turn on a lack of epistemic access to the possible referents of their respective moral terms. Each community is in fact in causal and even epistemic contact with instances of both properties.

17 Again, a point well made in Schroeter and Schroeter (ms.). I should add that the point is pretty general since many theories of reference require that we stand in the right epistemic relation to the things to which we refer. As David Copp helpfully noted to me in response to this chapter, "Could any term come to refer to a property in a way that did not depend on our ability to discriminate the property?"

18 I suppose one could be a Platonist about ethics but not about other properties, but I can't see why one would hold that view.

19 Such internalism falls immediately out of a view like Scanlon's, which just works with reasons as the basic normative notion. It is not as clear whether other prominent non-naturalists like Enoch or Shafer-Landau would endorse it. But still it seems to me as one of the better ways to try to cash out the alleged differentness of the normative, though other morals/reasons internalists are happy to embrace naturalism.

20 It is in fact striking that Street (2008: 214ff.) brings in a dilemma involving morals/reasons internalism to answer Copp's (2008) stance-independent realist response to her challenge. She suggests that only a reasons internalist reading of this theory is genuinely normative, but then argues that this kind of theory has trouble explaining congruence. If the theory is reasons externalist, it isn't normative, she suggests. That's in essence the same dialectic as is going on here for the non-naturalist. If your theory captures the motivations that make internalism attractive, you have trouble retaining a stance-independent view.

21 Or more weakly, any property that had (for example) racism or some other intrinsically wrong practice in its extension could not give us reason to go in for such things, in the way that the rightness of anti-racism gives us reason to be against racism.
22 Many thanks to Harry Ide and Joe Mendola for helpful discussion of the ideas in this chapter, to Matt Bedke for help with the relevant literature, and to David Copp, Diego Machuca and Folke Tersman for very helpful comments on a draft of this chapter.

References

Bedke, M. 2009. "Intuitive Non-Naturalism Meets Cosmic Coincidence," *Pacific Philosophical Quarterly* 90: 188–209.
———. 2014. "No Coincidence?" In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Volume 9, 102–125. Oxford: Oxford University Press.
Boyd, R. 1988. "How to Be a Moral Realist." In G. Sayre-McCord (ed.), *Essays on Moral Realism*, 181–228. Ithaca: Cornell University Press.
Brink, D. 1989. "Externalist Moral Realism," *Southern Journal of Philosophy* 24 (supplement): 23–40.
Brosnan, K. 2011. "Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge?" *Biology and Philosophy* 26: 51–64.
Clarke-Doane, J. 2012. "Morality and Mathematics: The Evolutionary Challenge," *Ethics* 122: 313–340.
———. 2014. "Moral Epistemology: The Mathematics Analogy," *Noûs* 48: 238–255.
Copp, D. 2008. "Darwinian Skepticism about Moral Realism," *Philosophical Issues* 18: 186–206.
Dowell, J. L. 2016. "The Metaethical Insignificance of Moral Twin Earth." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Volume 11, 1–27. Oxford: Oxford University Press.
Dreier, J. 1990. "Internalism and Speaker Relativism," *Ethics* 101: 6–26.
Enoch, D. 2010. "The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It," *Philosophical Studies* 148: 413–438.
———. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.
FitzPatrick, W. 2015. "Debunking Evolutionary Debunking of Ethical Realism," *Philosophical Studies* 172: 883–904.
Hilbert, D. 1987. *Color and Color Perception*. Palo Alto: CLSI.
Horgan, T., & Timmons, M. 1992. "Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived," *Philosophical Papers* 21: 153–175.
Joyce, R. 2006. *The Evolution of Morality*. Cambridge: MIT Press.
———. 2016. "Reply: Confessions of a Modest Debunker." In U. Leibowitz & N. Sinclair (eds.), *Explanation in Ethics and Mathematics*, 124–145. Oxford: Oxford University Press.
McDowell, J. 1985. "Values and Secondary Qualities." In T. Honderich (ed.), *Morality and Objectivity*, 110–129. London: Routledge.
Milo, R. 1995. "Contractarian Constructivism," *Journal of Philosophy* 92: 181–204.
Moore, G. E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.
Pettit, P. 2015. "The Birth of Ethics," *Tanner Lectures*, delivered at the University of California Berkeley.
Railton, P. 1993. "What the Non-Cognitivist Helps Us to See the Naturalist Must Help Us to Explain." In J. Haldane & C. Wright (eds.), *Reality, Representation, and Projection*, 279–300. Oxford: Oxford University Press.

Scanlon, T. M. 2014. *Being Realistic about Reasons*. Oxford: Oxford University Press.

Schafer, K. 2010. "Evolution and Normative Scepticism," *Australasian Journal of Philosophy* 88: 471–488.

Schroeter, L., & Schroeter, F. 2014. "Normative Concepts: A Connectedness Model," *Philosophers' Imprint* 14 (25): 1–26.

———. (ms). "Can Metasemantics Resolve the Epistemic Challenge to Moral Realism?" www.academia.edu/2923803/Can_Metasemantics_Resolve_the_Epistemic_Challenge_to_Moral_Realism. (Accessed 1/1/2017).

Shafer-Landau, R. 2003. *Moral Realism*. Oxford: Oxford University Press.

———. 2012. "Evolutionary Debunking, Moral Realism and Moral Knowledge," *Journal of Ethics and Social Philosophy* 71: 1–37.

Skarsaune, K. O. 2011. "Darwin and Moral Realism: Survival of the Iffiest," *Philosophical Studies* 152: 229–243.

Street, S. 2006. "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127: 109–166.

———. 2008. "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying about," *Philosophical Issues* 18: 207–228.

———. 2011. "Mind-Independence without the Mystery: Why Quasi-Realists Can't Have It Both Ways." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Volume 6, 1–32. Oxford: Oxford University Press.

Sturgeon, N. L. 2006. "Ethical Naturalism." In D. Copp (ed.), *The Oxford Handbook of Ethical Theory*, 91–121. Oxford: Oxford University Press.

Tropman, E. 2014. "Evolutionary Debunking Arguments: Moral Realism, Constructivism, and Explaining Moral Knowledge," *Philosophical Explorations* 17: 126–140.

van Roojen, M. 2010. "Moral Rationalism and Rational Amoralism," *Ethics* 120: 495–125.

———. 2006. "Knowing Enough To Disagree: A New Response to the Moral Twin Earth Argument." In R. Shafer-Landau (ed.), *Oxford Studies In Metaethics*, Volume 1, 161–193. Oxford: Oxford University Press.

———. Forthcoming. "Rationalist Metaphysics, Semantics and Metasemantics." In K. Jones & F. Schroeter (eds.), *The Many Moral Rationalisms*. Oxford: Oxford University Press.

Vavova, K. 2014. "Debunking Evolutionary Debunking." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Volume 9, 76–101. Oxford: Oxford University Press.

Wielenberg, E. J. 2010. "On the Evolutionary Debunking of Morality," *Ethics* 120: 441–464.

9 Moral Skepticism and the Benacerraf Challenge

Folke Tersman

1. Introduction

Platonists in the philosophy of mathematics believe that mathematical beliefs posit objects (numbers, functions, sets, etc.) that are mind- and language-independent, causally inert, and exist beyond time and space. The Benacerraf challenge concerns how to reconcile that belief with the existence of mathematical knowledge. After all, if numbers belong to a realm that is entirely cut off from us in the indicated ways, how could we possibly have any access to the truth about their features and relations? Some suspect that there is no viable way of answering that question, which in turn is taken to be a reason to abandon Platonism and to adopt some competing view, such as a constructivist account of mathematical entities.

There are other forms of realism, in other philosophical domains, that make similar claims about the facts and properties they posit as those Platonists make about numbers. Those positions are therefore held to be vulnerable to analogous challenges. One candidate is ethical non-naturalism. That ethical non-naturalists, and moral realists in general, face difficulties in accounting for the existence of knowledge in the target area is the conclusion of several well-known arguments. This is the upshot of some so-called "evolutionary debunking arguments" (see, e.g., Street 2006; Joyce 2013), but also of certain versions of the argument from moral disagreement (see, e.g., Tersman 2006: ch. 4), as well as of Gilbert Harman's contention that we have no reason to believe in objective moral facts because they are never assumed by the best explanation of anything observable (Harman 1977: ch. 1). However, while many of those more familiar arguments appeal to complex empirical assumptions, the point of the Benacerraf challenge is to show that the assumptions the target position makes about the facts it posits *by themselves* rule out the existence of knowledge. By applying that reasoning to ethics, the critics of non-naturalist moral realism may thus hope to avoid criticism of the kind the invocation of controversial empirical claims inevitably invites and to reach their conclusion in a more comfortable way.

The aim of this essay is to explore whether the Benacerraf challenge really does provide the critics of non-naturalist moral realism with such

extra resources. Is there a compelling version of the Benacerraf challenge that pertains to ethical non-naturalism and that is distinct from the more familiar ones just mentioned? That's the central question in this essay. The upshot is that this question should be answered in the negative. The moral version of the Benacerraf challenge must (unlike the mathematical one) be supplemented with considerations that are external to it, and does not ultimately constitute a distinct and separate objection that is independent of the more familiar ones.

2.   Field's Version of the Benacerraf Challenge

The Benacerraf challenge owes its name to Paul Benacerraf, who gave a first formulation of it in his seminal paper "Mathematical Truth" (Benacerraf 1973). Benacerraf's own version is widely seen as unsatisfying, however, because he explicitly presupposes a causal theory of knowledge whose plausibility can be questioned on independent grounds. I shall therefore focus on the more recent version that is due to Hartry Field.

Unlike Benacerraf, Field does not invoke the concept of knowledge. Instead, he takes the challenge to concern the "reliability" of our mathematical beliefs. According to Field, the problem with the claims Platonists make about the nature of mathematical entities is that they seem to exclude any viable explanation of the reliability of those beliefs. Field assumes that to explain the reliability of our beliefs in an area is to explain an "actual correlation" between our beliefs in that area and the relevant facts. That is, with regard to mathematical beliefs, it is to explain that the following schema

   If we accept 'p' then p

"(and a partial but hard to state converse of it) holds in nearly all instances, when 'p', in this case, is replaced by a mathematical sentence" (Field 1989: 26).[1]

In a sense, then, to explain the reliability of our mathematical beliefs is to explain why they are mostly true. Or, better, it is to explain why we have managed to arrive at beliefs in this area that are mostly true, given that this is indeed the case, since it is not the truth of their *contents* that is to be explained.[2] Field sometimes formulates his challenge a bit differently. What is to be explained (or shown by our account) is that, assuming that we have in fact arrived at mathematical beliefs that are mostly true, this is not a result of a "huge coincidence" or sheer luck.

Now, on Field's construal of the Benacerraf challenge, a central premise is the claim that the principled inability to provide such an explanation tends to undermine the justification or epistemic permissibility of those beliefs.[3] The underlying intuition is illustrated by Field's famous story about a remote village in Nepal. Suppose that we initially have various beliefs

about this village, such as that an odd number of its inhabitants are born on a Friday, perhaps because we have been told so by someone who claims to have lived there. Suppose also that we subsequently learn that the person in question has never in fact been in any contact with the village, and that there is no alternative account of the reliability of our beliefs that holds any promise. In that case, it would be wise of us to drop the beliefs and to adopt an agnostic attitude towards claims about the goings-on in the village. The analogy is stressed in the following passage:

> Consider the Benacerraf worry about (certain kinds of) platonism in the philosophy of mathematics. The worry is that the same reasons that would lead us to advise against having beliefs about the happenings in a remote village in Nepal, when one has reason to think that there is no possible explanation of the reliability of those beliefs, should equally lead us to advise against having beliefs about mathematical entities platonistically construed, given that it appears that there is no possible explanation of the reliability of those beliefs.
>
> (Field 2009: 289)

The quote suggests the following formulation of the epistemic principle that Field has in mind and that serves as a first premise of his argument:

A   If it is in principle impossible to explain the reliability of a set of beliefs, then they are unjustified and we should abandon them.[4]

The second step of the argument is the claim that Platonism excludes the possibility of a viable explanation of the reliability of our mathematical beliefs. Thus:

B   Given Platonism, it *is* in principle impossible to explain the reliability of our mathematical beliefs.[5]

And from **A** and **B** we can derive:

C   Given Platonism, we should abandon our mathematical beliefs.

Note that **C** does not by itself exclude the correctness of Platonism. So how is the argument supposed to proceed? Given what I said at the outset (that the problem highlighted by the Benacerraf challenge is taken to consist in the fact that Platonism cannot accommodate the existence of mathematical knowledge), it may seem tempting to construe it as a *reductio ad absurdum*. On that idea, **C** undermines Platonism because the conclusion that we should abandon our mathematical beliefs is so implausible and conflicts with our intuitions. Therefore, the fact that Platonism implies that our mathematical beliefs are not epistemically permissible represents a high

theoretical cost. It may perhaps have advantages that compensate for this cost. But in the absence of such advantages, it is a conclusive reason to reject Platonism.

Although this construal may square with how some commentators conceive of the argument, it does not capture Field's conception. For Field *agrees* with the conclusion that we should drop our mathematical beliefs, in the sense that we should cease to take the contents of our mathematical beliefs to be "literally true." In fact, that's the central component of his "fictionalist" view on the nature of mathematics. Field's point is rather that, by ceasing to take the contents of the beliefs to be true, we are no longer Platonists. Note that Platonism consists of two elements, a "conceptual" one and a "metaphysical" one. That is, to be a Platonist is to think that our mathematical beliefs posit entities that exist beyond time and space, etc. But it is also to *have* some of the beliefs Field thinks should be given up and to believe that those entities actually exist. What the Benacerraf challenge potentially brings out is a tension between those elements. The idea is that the Platonist, by making assumptions that rule out the possibility of a viable explanation of the reliability of our mathematical beliefs, commits a kind of epistemic suicide. For what follows from **C** is, in effect, that, given Platonism, we should not be Platonists.

This implication does not establish the falsity of Platonism, of course. After all, the fact that a position implies about itself that it is unjustified is not inconsistent with its truth. However, it does seem to undermine Platonism in the sense that it supports not only the hypothetical claim that *given* Platonism, we should not be Platonists, but also the categorical claim that we should not be Platonists full stop. Of course, that conclusion does not follow either, strictly speaking. But if we are forced to concede that a position we have adopted is permissible only if it is false, then we have surely ended up in an awkward place. In any case, on the second, alternative construal of the argument, it continues in the following way:

D Given Platonism, we should not be Platonists.
E Therefore, we should not be Platonists.

We may accordingly distinguish between two versions of the argument that the Benacerraf challenge potentially generates (something I hereafter will call "the BF-argument"). On the first version, the alleged inability to explain the reliability of our mathematical beliefs on Platonist grounds provides a reason to think that Platonism is false, since it means that it conflicts with an important intuition. On the second version, it rather threatens Platonism by rendering it epistemically self-defeating (in the way indicated by **D**). This may plausibly be taken to undermine the justification with which the position is held even if it doesn't undermine it in the more straightforward sense of showing it to be false. We may call the first version the "intuitive version" and the second "the self-defeat version."

3. Platonism and Non-Naturalism

Whether a plausible and distinct argument that is parallel to the BF-argument can plausibly be mounted against ethical non-naturalism depends on a number of issues. It depends on how closely related the positions are. But it also depends, as I shall illustrate, on which responses are available to an ethical non-naturalist and to a Platonist.

Ethical non-naturalists believe in the existence of mind- and language-independent moral facts and properties. What makes their position non-naturalist is that they deny that those facts and properties are reducible to natural ones. Which facts are natural? Answering this question in an enlightening way has turned out to be difficult, but some are sympathetic to Michael Smith's suggestion that it depends on whether the fact is "the subject matter of a natural or social science" (Smith 1994: 17). Ethical non-naturalism is also often associated with the thesis that the moral facts and properties it posits are causally inert.[6]

Now, Platonism makes similar assumptions about mathematical facts. But, as the above account indicates, what is often supposed to give rise to the Benacerraf challenge is rather the fact that Platonists in addition make those assumptions about the *objects* mathematical terms are supposed to denote, such as numbers. Those entities are also taken to not bear any causal or spatio-temporal relations with us. Ethical non-naturalists, by contrast, don't posit any objects of that type. The bearers of moral properties are straightforwardly natural phenomena, such as actions, political institutions, or people, and no assumption is made to the effect that we don't causally interact with those phenomena. It may perhaps be argued that this makes a difference, and that it makes non-naturalists less vulnerable to the BF-challenge. However, that is not a strategy that I shall pursue.

What I do want to argue is that there are responses to the versions of the BF-argument that applies to morality that do not seem available to a Platonist about mathematics. To block those responses, the moral BF-argument must be supplemented with considerations that are external to the argument.

The responses in question consist in developing so-called "third-factor" explanations of the reliability of our moral beliefs. Consider Enoch's (2010) version of this strategy.[7] As we have seen, to explain the reliability of our moral beliefs is to explain an alleged correlation between our moral beliefs and the facts in virtue of which they are supposed to be true. Now, if the facts belonging to one set (*a*-facts) are correlated with those that belong to another set (*b*-facts), we may try to explain this either by arguing that *a*-facts cause *b*-facts or vice versa. However, the idea behind the third-factor strategy is instead to point to an external factor that is supposed to be responsible for both types of facts.

Let us focus in what follows on "rightness beliefs," i.e., beliefs to the effect that certain types of actions are morally right. Enoch tries to provide the sought-for explanation by invoking an evolutionary account according

to which we are disposed to have the rightness beliefs we have because the actions they prescribe increase the agent's fitness. According to the version of the evolutionary account Enoch has in mind, moreover, they often do so through helping the agent or her kin to survive. It is that fact—the fact that the actions in question promote survival—which constitutes the third factor to which Enoch appeals.

How does the explanation work? An important element is a further assumption that Enoch invokes, namely the substantively moral claim that "survival or reproductive success [. . .] is at least somewhat good" (2010: 430). The claim is clarified as follows:

> Not, of course, that it is always good, or that its positive value is never outweighed by other considerations, or even that it is of ultimate or of intrinsic value, or anything of the sort. Furthermore, I am not asking you to assume that the evolutionary "aim" is of value *because* it is the evolutionary aim. All I will be relying on is the assumption that survival (or whatever) is actually by-and-large better than the alternative.
>
> (2010: 430)

Given this assumption, we may reason as follows. On the evolutionary account, the factor that explains our rightness beliefs is the fact that the actions they ascribe rightness to tend to promote survival. But, given that survival is good, this factor also tends to make the actions right, at least if we make the additional assumption that the fact that an act promotes the good counts in favor of its rightness. Thus, the same factor that is supposed to be responsible for the existence of our rightness beliefs—that the actions they prescribe promote survival—is also, at least partially, responsible for their truth. Enoch thinks that his account suggests that many of the beliefs we are endowed with by evolution are at least "somewhat in line with the normative truths" (2010: 430).[8] He concedes that the level of reliability we can squeeze out is not overwhelming, as he thinks it merely suggests that the moral beliefs we have been endowed with through evolution are "reasonably good starting-points" that are not "too far off." But he thinks that it still gives non-naturalists what they need.

The salient feature of Enoch's proposal is that it tries to reconcile the reliability of our moral beliefs with the evolutionary account claim by invoking a substantive moral assumption. The assumption in question posits a relation between the facts in virtue of which our moral beliefs are true and the (non-moral) facts that, according to the evolutionary account, explains why we have those beliefs.

There is an obvious way of criticizing Enoch's strategy, namely by denying that he is entitled to invoke such a bridge principle. After all, the viability of the response arguably presupposes that he has some reason to *accept* the bridge principle in question. And since the principle in question is a substantive moral claim, the assumption that there is such a reason is in direct

conflict with the conclusion advocates of the moral BF-argument think non-naturalists are committed to, namely that no moral belief is justified. In other words, by invoking the pertinent bridge principle, Enoch seems to blatantly beg the question.

However, in my view, this objection is not persuasive. For by requiring that a non-naturalist be able to justify the reliability of moral beliefs in an entirely non-question-begging way is to impose too strong a condition that risks collapsing into global skepticism. For example, consider perceptual beliefs. We do have an account of the reliability of such beliefs; an account that consists of appeals to the theory of evolution and to assumptions about how our sense organs work and how they interact with our environments, and so on. This account involves the same circularity as Enoch's does. For the assumptions on which it rests constitute straightforward empirical claims whose justification depends on the credentials of the very beliefs whose reliability it tries to establish. So if we deem Enoch's reasoning illegitimate on account of its circularity, we seem committed to the same verdict about our perceptual beliefs. Accordingly, insofar as the advocates of the BF-argument don't want their position to collapse into some sort of global skepticism, the circularity that is involved by Enoch's strategy should be deemed legitimate.

Enoch's response to the BF-argument could obviously be questioned in various other ways. However, we may imagine further instances of his strategy, and the points I presently want to make do not presuppose the soundness of that particular version.

The first point is simply that the same type of response does not seem available to a Platonist about mathematics. It is not available since Platonists are not in any position to argue that the natural facts that are causally responsible for our mathematical beliefs are also (metaphysically) responsible for mathematical facts. This is important in the present context, because (and this is my second point) an advocate of the moral BF-argument will, unlike an advocate of the mathematical one, need some way of ruling out responses of that sort.

How could they be ruled out? It is at this juncture where the more familiar arguments to the effect that non-naturalism entails skepticism have a potential role to play.

The point of departure of one critical strategy is the aforementioned observation that the degree of reliability of the target judgments that one might hope to secure through Enoch's account is rather low. Darwinian factors might have ensured that our rightness beliefs track types of behavior that promote our *own* chances of survival or of the survival of *our* kin. But that doesn't necessarily mean that they are similarly sensitive to facts about how those types of behavior affect the survival of *others*. That detracts from their reliability. For if the claim that survival is good is supposed to concern survival in general, the effects on others are also relevant to the rightness of the target behavior. Moreover, survival is presumably not the only thing that is good and whose existence can influence the rightness of that behavior.

So the forces of natural selection have at best ensured that our rightness beliefs are sensitive to a very limited set of the considerations that determine rightness.

This is a potential problem. For to provide a response to the skeptical challenges, a realist must not only reconcile her position with the existence of knowledge about trivialities (such as that pain is bad or that survival is "somewhat" good), but also about more controversial issues, including, for example, what weight one is permitted to assign to one's own survival when facing a situation where the survival of many others is also at stake. It is hard to see how one could reach justified conclusions about such issues merely on the basis of the account Enoch sketches and the level of reliability it allegedly ensures.

To strengthen the case for the significance of the account, Enoch points to the additional, benign effects of ordinary reasoning processes. He writes:

> Given a starting point of normative beliefs that are not too far-off, presumably some reasoning mechanisms (and perhaps some other mechanisms as well) can get us increasingly closer to the truth by eliminating inconsistencies, increasing overall coherence, eliminating arbitrary distinctions, drawing analogies, ruling out initially justified beliefs whose justificatory status has been defeated later on, etc.
>
> (2010: 428)

It is the *combined* effect of these factors—the fact that our initial beliefs about what is right and wrong depend on the extent to which they promote survival and the benign effects of subjecting them to a reflective equilibrium-style reasoning process—that are thought to be sufficient for the non-naturalist to avoid skeptical implications.

However, it is not clear how much "extra" reliability the influence of this type of process really can generate. Presumably, all inquirers are affected by the genetic dispositions we have inherited through evolution, although some may be more capable of distancing themselves from them than others. Still, the history of moral inquiry illustrates that different individuals, when seeking coherence in the way Enoch suggests, end up with radically different moral outlooks. Some turn out as consequentialists whereas others become virtue theorists or particularists, and this is so even if we lack compelling (and independent) grounds for thinking that some of them are less well equipped with the relevant reasoning skills than others. Disagreement of that type provides a potential problem for strong claims about the reliability of the process Enoch describes in the quote. Accordingly, it also potentially undermines the idea that one, by appealing to it, can strengthen the case for the viability of Enoch's response to skeptical challenges. We seem thus to be back at square one.

Notice that my aim here is not to try to refute Enoch's response through the above remarks. The point is rather to illustrate that in fending off

Enoch-style responses, an advocate of the moral version of the BF-argument must invoke considerations that are external to the argument, such as disagreement. That should dampen our expectations about the independence of the support it can provide for the anti-realist's position.

4. Safety

We get further reasons for this conclusion by considering another recent response to the BF-argument, namely a response that, unlike Enoch's, *is* available to the Platonist.

The response in question has been articulated by Justin Clarke-Doane (2016). His idea is that, to put the skeptical worries that might be raised by our inability to explain the reliability of a set of target beliefs to rest, it is sufficient to provide a compelling argument to the effect that, given the truth of the beliefs, they could not easily have been false (in which case they are said by Clarke-Doane to be "safe"). And his point is that a Platonist *can* consistently construct such an argument, in spite of her peculiar assumptions about the nature of mathematical objects.

The argument proceeds as follows. One way in which our beliefs in an area could easily have been false is through being such that they would have remained the same (had the same contents) even if the facts they represent had been relevantly different. Consider a bathroom scale that has a similar feature. Thus, suppose that it shows that a person weights 65 kg, but that we learn that it would have shown this even if the person had been significantly lighter. This undermines our reasons for trusting it, regardless of whether the person actually does weigh 65 kg. Now, according to Clarke-Doane, the fact that our mathematical beliefs are not unsafe in the indicated sense is ensured by the metaphysical necessity of mathematical facts. For their necessity ensures (on one view on the truth conditions of counterfactuals) that it is vacuously true that if the mathematical facts had been different, then our beliefs would have been correspondingly different. After all, if there is no possible world in which they are different, there is also no world in which they are different while our beliefs remain the same.

Given the necessity and actual truth of our mathematical beliefs, there is only one remaining way in which they could easily have been false. This is to argue that the beliefs could easily have been *different*, even if the facts had remained the same. For if we could easily have had different mathematical beliefs, then we could easily have had false mathematical beliefs, even supposing that they are in fact (necessarily) true. However, this alternative way of questioning the safety of our mathematical beliefs can also be ruled out, according to Clarke-Doane, at least in the cases of our "core" or most basic mathematical beliefs. It can be ruled out, more specifically, with reference to an evolutionary account of their origins. The idea seems to be that, given such an account, we would have had different beliefs only if our environment had been quite different, which means that it excludes the

existence of nearby possible worlds in which our (core) mathematical beliefs are different.

Clarke-Doane concludes that the necessity of mathematical facts and the evolutionary inevitability of our core mathematical beliefs explain their reliability in the sense of showing that, given their truth, they could not easily have been false. He also thinks that this is the only sense in which the inability to produce such an explanation undermines their justification. Moreover, since the explanation is consistent with the assumptions a Platonist makes about the nature of mathematical objects, it is clearly available to her. The same reasoning applies, according to Clarke-Doane, to ethical non-naturalism. For moral beliefs can also be given an evolutionary explanation and (some) moral facts also seem to hold with necessity. Neither the moral nor the mathematical version of the BF-argument therefore gets off the ground.

Let us grant Clarke-Doane the conclusion that Platonists and non-naturalists can consistently and plausibly argue that the target beliefs are safe on the basis of the considerations he adduces. In order for the BF-argument to survive this concession, there has to be an alternative interpretation of "explain reliability" to the one Clarke-Doane relies on. This alternative interpretation must in turn meet two conditions. It must, first, be such that the argument Clarke-Doane has proposed does not qualify as an explanation in that sense, and, second, such that it is plausible to hold that the principled inability to explain the reliability of the target beliefs in that sense undermines their justification.

5.   Explaining Reliability

Is there an interpretation that meets those conditions? We may begin to explore this question by reconsidering why the inability to explain the reliability of a set of beliefs is supposed to undermine them. Notice in this context that there are two ways to undermine a belief's justification. The most straightforward way is to offer some compelling reason to think that it is false. Another is to show that we don't have sufficient reason to think that it is *true*. If a consideration of the latter type does not also provide evidence for the falsity of the belief, it supports agnosticism with regard to its content rather than outright denial.

The inability to viably explain the reliability of a set of beliefs, such as the set that includes our beliefs about the village in Nepal, seems at best to undermine them in the second way. After all, the fact that we cannot construct such an explanation hardly shows that an odd number of the inhabitants are *not* born on a Friday. Let us therefore assume that insofar as the inability to explain the reliability of our beliefs in some area undermines them, it does so through ruling out that we have sufficient *positive* reasons for believing in their truth.

If we want to reconstruct the BF-challenge along those lines, we could do so by arguing that the availability of a compelling reliability-explanation

would itself be (or provide) a positive reason for the target beliefs. For then we could go on to claim the *absence* of such an explanation deprives the beliefs of precisely that reason.

Note in this context that the explanation Clarke-Doane offers (i.e., the one that appeals to the necessity of mathematical and moral facts and the evolutionary inevitability of our beliefs about them) does *not* constitute (or provide) a positive reason for them. What a Platonist or non-naturalist might accomplish by citing those considerations is to show that, *if* the beliefs are mostly true, then they could not easily have been false. But that conclusion does not in turn give us any reason to conclude that they *are* in fact true. Consider an analogy. Suppose that two friends visited a certain city in the weekend and that we wonder if they met. Suppose also that we know that each had a very determinate plan about her visit. Then we may perhaps conclude that if they met, this was not a coincidence and that it could not easily have been different. But we would still be in the dark as to whether they actually *did* meet.[9] Clarke-Doane's account of the reliability of our mathematical beliefs does accordingly not serve the dialectical role the BF-argument assigns to such accounts, given the reconstruction of it that we are now pondering.

Which types of explanations, then, *do* serve that role? I wrote earlier that to explain the reliability of our beliefs is to explain that we have arrived at beliefs in the relevant area which are mostly true. Presumably, this involves explaining why we *have* arrived at those beliefs in the first place. Some such explanations are neutral relative to the truth of the beliefs, in the sense that the truth of the assumptions they involve doesn't affect or increase the beliefs' likelihood of being true. For example, the fact that we have formed them as a result of reading a book is thus neutral as long as we don't know anything about its trustworthiness. Other explanations *do* increase their likelihood, however, by indicating or providing grounds for thinking that they are true. Indeed, the best explanations of some of our beliefs even *presuppose* their truth, such as when we explain why we have the perceptual belief that there is a table in front of us by citing the fact that there *is* a table in front of us. On one suggestion, it is explanations of the second kind—explanations of why we have the target beliefs that *also* provide grounds for thinking that they are true—that we should refer to as "explanations of reliability." This understanding of what constitutes an explanation of the reliability of a set of beliefs fits, unlike Clarke-Doane's, with the interpretation of the BF-argument we are presently considering. For if the best explanation of a set of beliefs provides a reason for thinking that they are true, we may argue that the inability to provide such an explanation undermines their justification by depriving them of that reason (premise **A**).

A problem with this suggestion is that the support it provides for **A** seems very weak. After all, the fact that *one* potential source of justification for a belief is ruled out may not exclude that there are other sources. Consider again our beliefs about the village in Nepal and suppose that the

best explanation of why we have them does not indicate that their contents are true. This is consistent with the fact that they are related in such a way that each of them gets some conditional support from the others, so that, for example, our belief that an odd number of inhabitants are born on a Friday can be backed up with our belief that Sonam, Kiran, and Yash were born on a Friday and that no one else lives in the village. The same holds for our mathematical beliefs. Why isn't that type of internal coherence enough for justification or epistemic permissibility?

The answer must presumably be that there is something special about the support provided by explanations of reliability that makes their absence especially significant. One option is to stress that it is external to, and/or independent of, the (the truth of the) target beliefs. If the best explanation of our having the belief that P implies or strongly indicates the truth of P, then the fact that we have it is evidence for P, regardless of the content of P. The fact that we have the belief is a psychological and natural fact, and can be established without settling any moral or mathematical issues. This is why—on one idea—the availability of an explanation of the reliability of our moral or mathematical beliefs provides crucial support for them and why the principled absence of such an explanation may have skeptical implications.

The reasoning just hinted at needs to be extensively elaborated to have any prospects of success, for example regarding the concept of "independence" that might plausibly be invoked. Let me again stress, however, that my aim in this essay is not to defend any particular construal of the BF-argument. The point I want to make is rather that it makes the request to show how the reliability of moral or mathematical beliefs could possibly be explained on non-naturalist or Platonist grounds collapse into the more familiar demand to show how the permissibility of our beliefs in the target areas can be reconciled with a naturalistic epistemology.

Note that, if it is the naturalness of the fact that we have a belief that is the source of the significance the BF-argument assigns to explanations of reliability, then support from *any* kind of natural facts would do just as well. For example, a Platonist could try to respond to the argument by stressing, à la Quine, that (at least some subset of) our mathematical beliefs are justified in virtue of playing an indispensable role in physics. And a moral realist may hope to do the same with regard to morality by appealing to the role moral assumptions possibly play in the explanations of phenomena such as, to use Nicholas Sturgeon's (1988) example, social revolutions. Of course, both these defensive projects might fail. But the thing to note is that if an argument to the effect that the best explanation of some natural phenomenon assumes moral facts counts as a response to the moral version of the BF-argument, then it coincides with Harman's argument to the effect that we have no reason to believe in the existence of objective moral facts because they are never posited by the best explanation of anything observable.

Hartry Field is of course associated with the view that the Quinian argument *does* fail, through his attempts to "nominalize" physics and to show

that the role mathematics play isn't indispensable after all.[10] But he has also suggested that the outcome of that discussion is really irrelevant to the Platonist's ability to handle the BF-challenge and that a successful indispensability argument would not really help the Platonist against it. One may accordingly conclude that the construal of the argument developed in this section fails to capture Field's intentions. However, Clarke-Doane's argumentation illustrates that the room for an alternative construal is slight. This provides further support for the thesis that, insofar as the BF-argument has any weight at all, it does not constitute a distinctive, separate objection.[11]

6. Concluding Remarks

We started out by wondering if the Benacerraf challenge allows the critics of moral non-naturalism to expand their dialectical arsenal, by offering a way of reaching their conclusions without having to get their hands dirty through empirical speculations. The upshot of my discussion is that it doesn't really. When we seek to reconstruct the challenge in a way that avoids crucial objections, it is transformed into a version of Harman's well-known debunking-style argument and must be supplemented, for example, with considerations invoked by the argument from disagreement.

I would like to end, however, with some partially self-critical remarks. The aim to identify distinct arguments for a philosophical position is understandable because it may enable us to proceed in a piecemeal fashion when we want to assess it. We could start by looking at one argument, and when we're done we turn to the next, and so on, until an overall evaluation is reached. In reality, however, many arguments in fact interact in different ways, and the piecemeal strategy sometimes leads us to ignore the existing dependencies, which in turn makes us less well placed to assess the argument. This observation calls, I think, for a more holistic approach. As John Rawls stressed, the justification of a philosophical position is "a matter of the mutual support of many considerations, of everything fitting together into one coherent view" (Rawls 1971: 21). It may seem that I have violated this dictum by focusing on the question of whether the moral BF-argument is a distinct objection. However, the conclusion I have defended can be formulated in a way that acknowledges it. The conclusion is simply that the BF-argument doesn't contribute much to the relevant whole. It relies on basically the same epistemic ideas and machinery as the more familiar challenges but adds very little, if anything, to the bigger picture.

Notes

1 Field focuses on the beliefs held by *mathematicians* and not those that are held by people in general ("us"), but that makes no difference here. I shall continue to write about "our" mathematical beliefs.

2 By a "belief" I refer to a certain psychological state or an instance of a certain psychological attitude. Each such state has a content that is constituted by a

proposition, or so I assume, and people are said to "share" beliefs to the extent that they have beliefs with the same contents.

3 At one juncture, he writes that "a principled inability to [explain the reliability of some set of belief] tends to undermine the justification of those beliefs" (1996: 377).

4 In some passages, Field formulates the antecedent of the principle so that it merely implies that we have *some* reason to give up the beliefs, such as when he writes that the absence of an explanation of the relevant kind "tends" to undermine the justification of the target beliefs. Field also offers formulations given which it is the "appearance" of the impossibility of providing such an explanation, or our possession of reasons for thinking that this is impossible, that are the crucial considerations rather than whether it *in fact* is impossible (Field 1989: esp. 25–30, 230–239; 1996; 2005). These variations do not, however, have any bearing on the arguments I shall pursue.

5 Note, however, that although Field is pessimistic about the possibility of explaining the reliability of our mathematical beliefs on Platonists grounds, he never really commits to **B**.

6 The causal inertness is assumed by David Enoch's (2011) version of non-naturalism. However, other non-naturalists, such as Russ Shafer-Landau (2003), seem less committed to it.

7 For related suggestions, see Schafer (2010), Wielenberg (2010), Brosnan (2011), and Skarsaune (2011).

8 A central concept in Enoch's account is that of a fact being "responsible" for some other fact. Enoch takes this term to cover both a causal relation and "a metaphysical relation of constitution." It is the latter type of relation that is supposed to hold between the fact that an action promotes survival and the fact that it is right. Note, however, that, as Enoch rejects naturalism, the relation is not that of identity. For more about the type of relation Enoch has in mind, see Enoch (2010: 431).

9 If we were better informed about the details of their plans, then we would perhaps be able to figure out if they met. But the point is that this further information is not required to reach the conclusion that *if* they met, it could not easily have been otherwise.

10 See Field (1980). For a defense of mathematical indispensability arguments, see Colyvan (2001).

11 For further discussion of Clarke-Doane's arguments and their implications for skeptical challenges against moral realism, see Tersman (2016).

References

Benacerraf, P. 1973. "Mathematical Truth," *Journal of Philosophy* 70: 661–679.

Brosnan, K. 2011. "Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge?" *Biology and Philosophy* 26: 51–64.

Clarke-Doane, J. 2016. "What Is the Benacerraf Problem?" In F. Pataut (ed.), *New Perspectives on the Philosophy of Paul Benacerraf: Truth, Objects, Infinity*, 17–43. Dordrecht: Springer.

Colyvan, M. 2001. *The Indispensability of Mathematics*. Oxford: Oxford University Press.

Enoch, D. 2010. "The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It," *Philosophical Studies* 148: 413–438.

———. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.

Field, H. 1980. *Science without Numbers*. Princeton: Princeton University Press.

———. 1989. *Realism, Mathematics, and Modality*. Oxford: Blackwell.

———. 1996. "The a Prioricity of Logic," *Proceedings of the Aristotelian Society* 96: 359–379.

———. 2005. "Recent Debates about the a Priori." In T. Gendler & J. Hawthorne (eds.), *Oxford Studies in Epistemology*, 69–88. Oxford: Oxford University Press.

———. 2009. "Epistemology without Metaphysics," *Philosophical Studies* 143: 249–290.

Harman, G. 1977. *The Nature of Morality*. New York: Oxford University Press.

Joyce, R. 2013. "The Evolutionary Debunking of Morality." In J. Feinberg & R. Shafer-Landau (eds.), *Reason and Responsibility: Readings in Some Basic Problems of Philosophy*, 15th edition, 527–534. Boston: Wadsworth, Cengage Learning.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Schafer, K. 2010. "Evolution and Normative Scepticism," *Australasian Journal of Philosophy* 88: 471–488.

Shafer-Landau, R. 2003. *Moral Realism: A Defence*. Oxford: Oxford University Press.

Skarsaune, K. 2011. "Darwin and Moral Realism: Survival of the Iffiest," *Philosophical Studies* 152: 229–243.

Smith, M. 1994. *The Moral Problem*. Oxford: Blackwell.

Street, S. 2006. "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127: 109–166.

Sturgeon, N. 1988. "Moral Explanations." In G. Sayre-McCord (ed.), *Essays on Moral Realism*, 229–255. Ithaca: Cornell University Press.

Tersman, F. 2006. *Moral Disagreement*. New York: Cambridge University Press.

———. 2016. "Explaining the Reliability of Moral Beliefs." In U. Leibowitz & N. Sinclair (eds.), *Ethics and Explanation*, 37–57. Oxford: Oxford University Press.

Wielenberg, E. J. 2010. "On the Evolutionary Debunking of Morality," *Ethics* 120: 441–464.

10 Veneer Theory

Aaron Zimmerman

> "As a nation, we began by declaring that, *'All men are created equal.'* We now practically read it 'All men are created equal, *except negroes.*' When the Know-Nothings get control, it will read 'All men are created equal, except negroes, and *foreigners, and Catholics.*' When it comes to this I should prefer emigrating to some country where they make no pretence of loving liberty—to Russia, for instance, where despotism can be taken pure, and without the base alloy of hypocrisy."
>
> (Lincoln 1855)

1. Introduction

Someone who means you harm pretends to be a friend. A *veneer* of goodwill effectively masks her self-serving motives. Her apparent concern is mock concern, and thus no concern at all.

When wielded as a slogan, "morality is a veneer" serves as an expression of moral skepticism. At a minimum, it suggests the infrequency of altruism and the commonality of injustice. At a maximum, it implies that moral discourse is a complete sham. We advise brotherly love and demand fairness and respect. But a person will neither give love nor lend respect unless we compel them or appeal to various ulterior ends. Morality is a veneer: a shiny coat of social grease applied to hide amoral selves too rough to interact without violent friction. Morality is mostly pretense: systemic deception.

In *Morality and the Social Instincts*—the Tanner Lectures on Human Values—the esteemed primatologist Frans de Waal coins the term "veneer theory" to describe views on which morality is a cultural construct. He therein converts "morality is a veneer" into something more than a skeptical slogan designed to draw our attention to commonplace forms of hypocrisy.[1] According to de Waal, T. H. Huxley (1895) was the paradigmatic veneer theorist.[2] Against Huxley, de Waal utilizes contemporary primatology to argue that the primate species from which the earliest humans, bonobos and chimps evolved had capacities for empathy and sympathy, a sense of fairness and gratitude, a non-derivative concern for the well-being of their kin, and a proclivity to experience anger at perceived injustices. Confronted with

the manifestation of these traits by non-human apes and other primates, we must conclude that a contemporary human's moral compass is largely a biological phenomenon, the outgrowth of a genetically inherited "core morality" that emerged long before the evolution of our species.

In de Waal's hands, "veneer theory" is defined with reference to academic theories of morality's origins and current status. De Waal dons the hat of an intellectual historian to argue that Huxley was a veneer theorist, and that the ideas that constituted Huxley's veneer theory shaped the subsequent course of biology to such an extent that veneer theory became a standard assumption among those working in the field. But de Waal makes no claims about the role Huxley's acceptance of veneer theory had on Huxley's politics nor on his interactions with family, friends and so on. Insofar as veneer theory is associated with the slogan "morality is a veneer," and that slogan is associated with the skeptical imagery evoked above, it is an interesting question whether this theory has "skeptical force" in either an epistemological or psychological sense. If someone is convinced of the academic view of morality's genesis and function that Huxley is supposed to have embraced, is she therein less likely to act in identifiably moral ways or less prone to judge people in moral terms? Does embracing Huxley's views of morality's origins and function somehow "commit" one to paying less attention to injustice, undeserved suffering and the rest of what one had previously conceptualized as pressing moral concerns?

As an autobiographical matter, I am skeptical of these theses when they are advanced in their most aggressive forms. No doubt, biologists are affected by what they read. And Huxley affected the moral views of his contemporaries to a greater degree than any biologist save Darwin. But genuine moral concern is not as fragile as the skeptics pretend. If Huxley's supposed assertion of veneer theory did undercut the prevalence and seriousness of the moral thoughts or actions of his audience, this was a slow process that has yet to flower in full.[3]

At any rate, we can suppose that the earliest humans had the biologically natural set of core capacities and proclivities to which de Waal points without forgetting the enormous diversity in moral codes reported by historians and anthropologists. Diverse tribes developed in diverse ways and warred with each other along the way. Group selection therein "winnowed" the diverse space of moral communities spawned by the initial tribe.[4] Some narrowing of the field is supposed to have happened before human language evolved (cf. Sterelny 2012). But by all accounts, the construction of sentential language proved crucial. Sentential language enabled people to formulate complex contracts, enter into complex economic arrangements and support religious, spiritual and political institutions to codify, inculcate and enforce those rules, norms and principles we associate with "morality" today.[5] These innovations bestowed tremendous advantages on communities competing and warring with groups of inarticulate animals—animals

that were limited to gestures, grunts, howls and chirps when communicating roles, duties and privileges to one another.

Still, while de Waal acknowledges that human language instituted a revolutionary alteration of primate morality, he argues that the roots of our contemporary norms in substantive capacities for fellow feeling, cooperative endeavor and norm-enforcement undermine the veneer theorist's analysis of moral talk in terms of pretense. Whatever one makes of the *metaphysics* of morals and the epistemological standing of moral judgments and experiences, morality is *biologically* and *psychologically real*. Since so much of morality existed before the evolution of human language (and thus human religion and human politics), morality today is much more than a story we tell to elicit behaviors that accord with our reproductive and pecuniary interests:

> We are not subduing the proverbial wolf within us or hypocritically fooling everyone around us when we act morally: we are taking decisions that flow from social instincts far older than our species, even though we add to these the perhaps uniquely human complexity of a disinterested concern for others and the society at large.
>
> (de Waal 2003: 33)

Clearly, de Waal finds some connection between the origins of our moral psychology and its current status; some relatively deep connection between (a) the innateness of those psychological faculties operative when we judge one another good or bad, fair or unfair, virtuous or vicious (and/or behave in ways that elicit these judgments); and (b) the biological reality (or "depth") of the traits we express or ascribe. After all, if de Waal didn't perceive a connection of this sort, he would not argue from the empathy, sympathy and sense of fairness exhibited by the chimpanzees he has studied to the optimistic conclusion that morality is much more than a veneer.

But is de Waal's inference cogent? Admittedly, discursive hypocrisy is uniquely human. Since the other primates cannot construct sentences, they cannot demand adherence to rules they secretly ignore. But the other primates aren't exactly saints. Though chimpanzees act benevolently and reconcile with one another after disputes in the ways de Waal so vividly describes, chimps are also prone to deception and domination (Byrne & Whitten 1988). Mightn't chimps engage in pretense with the aim of getting conspecifics to observe rules the pretenders have no intention of observing? And mightn't we have inherited from our primate ancestors a proclivity to engage in these tactics and a concurrent susceptibility to their deployment by others?

2.   Huxley's Theory of Morality

De Waal's attack on veneer theory has already been subjected to careful philosophical scrutiny. Indeed, the lecture series was reprinted along with reactions by R. Wright, C. Korsgaard, P. Kitcher and P. Singer in a volume

entitled *Primates and Philosophers* (de Waal 2006). But to my knowledge, "veneer theory" has not been situated within the taxonomies now standard among analytic meta-ethicists. So before assessing de Waal's case against veneer theory, we would do well to focus his reading of the target. For as de Waal conceptualizes it, veneer theory is a complex idea with at least four component theses. The first two theses concern the contemporary function of morality or the *uses* to which moral language and deliberate displays of apparent altruism and fairness are now put. The second two theses concern the *origins* of moral speech and our acts of seeming altruism and justice, and the biological and psychological robustness of these complexes. The first two theses characterize the use of moral language and display as deceptive and coercive; the second two theses limit morality to the effects of these deceptive, coercive acts. We can articulate these four claims as follows:

Morality is coercive: the main function of moral discourse and behavior is getting other people to behave in ways the speaker or actor desires, where the audience in question is not predisposed to act in the desired manner out of prudence or self-interest.

Morality is deceptive: moral discourse and behavior could not play their coercive function were this function advertised (or made explicitly known) to the audience at which it is aimed.

Morality is psychologically superficial: a typical human's default motives are immoral or amoral—genuinely moral behavior is invariably the upshot of the coercive and deceptive uses of moral language or behavior referenced above.

Morality is biologically superficial: most (if not all) of those aspects of our psychology that we owe to our evolved biology are either immoral or (at least) non-moral.

So understood, veneer theory is both non-cognitivist and error theoretic in orientation. Since the veneer theorist regards moral discourse as primarily coercive, she must conceptualize it as primarily prescriptive rather than descriptive. This is true of transparently prescriptive moral commandments and imperatives—"Honor your mother and father"—which we don't describe as true or false, accurate or inaccurate. But in arguing that our attempts to coerce one another constitute the "main function" of morality, veneer theory embraces the psychological analog of R. M. Hare's semantic prescriptivism, which goes further to analyze claims of right and permission, duty and obligation, virtue and vice, justice and fairness, goodness and badness in terms of injunctions and commands. Similarly, in arguing that moral language is typically deceptive, the veneer theorist embraces the psychological analog of J. L. Mackie's error theory. Surely, we wouldn't continue to issue moral commands were doing so entirely ineffectual. Since the veneer theorist claims that moral discourse could not play its coercive function were this function advertised (or made explicitly known) to the audience

at which it is aimed, she must think that the object of a moral command is typically deceived about the meaning, function or import of that command. The veener theorist insists with non-cognitivists of other stripes that moral language is ultimately grounded in command; and she joins error theorists in maintaining that we are massively ignorant of this fact, and mistaken too about the ends we are pursuing when issuing or obeying the commands in question.

Do those thinkers de Waal characterizes as veneer theorists embrace the four theses we have identified along with their skeptical consequences? Are these theses diagnostically adequate? At times, de Waal casts a wide net, calling Hobbes and several other theorists of the social contract "veneer theorists" because they posit "a rational decision by inherently asocial creatures" to explain the formation of civil societies (2003: 4).[6] But Hobbes is something of a straw man in this context. Social and political theorists have long rejected the idea of a social contract when that idea is put forth as a description of the emergence of civil societies from a supposed state of nature. (As the old saying goes, the social contract is not worth the paper it's *not* written on.) We don't need to employ contemporary primatology to flog a dead horse that we know can't respond to the beating.

Moreover, de Waal's main stalking horse is not Hobbes, but Darwin's bulldog, T. H. Huxley, who de Waal blames for the tendency of biologists and social scientists to adopt veneer theory as a working hypothesis (2003: 34). And Huxley is famous for delivering a particularly harsh *critique* of social contract theory. Indeed, after observing that Rousseau wisely deploys the social contract as an ideal rather than a description of mankind's past (1893a: 298–299), Huxley goes on to reject the ideal in question: "The political lantern of Rousseauism is a mere corpse candle and will plunge those who follow it in the deepest of anarchic bogs" (Huxley 1893a: 301).[7] If Huxley conceptualized morality as a veneer, it was not because he thought of it as a "choice."

So it is worth going back to Huxley, to see the degree to which Huxley embraced veneer theory's central theses as we've articulated them on de Waal's behalf. Did Huxley think of humans as "naturally" selfish animals? Did Huxley describe his contemporaries' use of moral language—or their more overt displays of apparent altruism and fairness—in terms of pretense? Did Huxley claim that moral assertions and displays of virtue primarily function to conceal the moralist's self-serving ends?

According to de Waal, Huxley inferred the biological superficiality of morality from a mistaken interpretation of natural selection. In a break with Darwin, Huxley is supposed to have ignored reciprocal altruism, kin selection and (non-kin) group selection, and because of these oversights he (i.e., Huxley) is supposed to have mistakenly concluded that evolution favors selfishness:

> Evolution favors animals that assist each other if by doing so they achieve long-term benefits of greater value than the benefits derived

from going it alone and competing with others. Unlike cooperation resting on simultaneous benefits to all parties involved (known as mutualism), reciprocity involves exchanged acts that, while beneficial to the recipient, are costly to the performer (Dugatkin 1997). This cost, which is generated because there is a time lag between giving and receiving, is eliminated as soon as a favor of equal value is returned to the performer. . . . It is in these theories that we find the germ of an evolutionary explanation that escaped Huxley.

(de Waal 2003: 10–11)

Why did evolutionary biology stray from this path during the final quarter of the previous century? This is probably due to the conviction of some prominent figures, inspired by Huxley, that there is no way natural selection could have produced anything other than nasty organisms. No good could possibly have come from such a blind process. This belief, however, represents a monumental confusion between process and outcome. Natural selection is indeed a merciless process of elimination, yet it has the capacity to produce an incredible range of organisms, from the most asocial and competitive to the kindest and gentlest. If we assume that the building blocks of morality are among its many products, as Darwin did, then morality, instead of being a human-made veneer, should be looked at as an integral part of our history as group-living animals, hence an extension of our primate social instincts.

(de Waal 2003: 34)

But did Huxley really deny the existence of group selection? Did he really overlook the possibility that competition between groups might select a population of cooperators?

No, he didn't.[8] Even the most cursory look at Huxley's writings will show that de Waal's charges are unjust on this score. Huxley explicitly hypothesized that competition between groups selected for what we might call "in-group" cooperativeness: the kind of *selective* goodwill manifested by those Europeans colonizing the New World.

There is no doubt of the result, if the work of the colonists be varied out energetically and with intelligent combination of all their forces. On the other hand, if they are slothful, stupid, or careless; or if they waste their *energies in contests with one another*, the chances are that the old state of nature will have the best of it. The native savage will destroy the immigrant civilized man. Of the English animals and plants some will be extirpated by their indigenous rivals, others will pass into the feral state and themselves become components of the state of nature. In a few decades, all other traces of the settlement will have vanished.

(Huxley 1895: 17, emphasis added)

Proof

There are two things to note about this initial quotation. The first is the *partial* nature of the cooperative motives that will emerge if group selection brings some change to the biological state of American nature. The distribution of phenotypes in the New World will remain largely unaltered unless the colonists cooperate with one another against the natives and other components of the natural order they found there. As we now know, the colonists did cooperate with one another, and did bring about a change in the distribution of phenotypes in the Americas. Like Darwin (1982 [1871]), Huxley viewed the history of colonization as a process of group selection.

So de Waal is wrong to accuse Huxley of ignoring every evolutionary mechanism save individual selection. It's just that Huxley didn't indulge in the absurd suggestion that group selection had yielded populations of pure altruists. Instead, competition for resources between groups of humans (and other animals) had left a highly *parochial* form of altruism in place: in-group solidarity rather than solidarity *simpliciter* (cf. Smith 2011). According to Huxley, the groups that had won in the battle for reproductive resources and therein persisted over time contained a higher proportion of what he called "cooperative intelligence." Holding all else equal, the members of less cooperative groups were outbred, extinguished or assimilated.[9]

Is cooperative intelligence, as Huxley conceived of it, a component of deep morality? In one sense of the question, the answer is "yes," but in another it's probably "no." Huxley thought of cooperative intelligence as a "deep" (biologically evolved) phenotype. But whether we conceive of the complex of psychological characteristics that explain a tribe's cooperation with one another as genuinely "moral" depends on the sense we lend to that essentially contested term. On Huxley's account, group selection yields the kind of narrowly trained moral ethos that was endorsed by the majority of his readers. Group selection had produced populations of individuals who were disposed to cooperate (more or less nicely) with those with whom they identified, but these were individuals who were equally disposed to war against those they conceptualized as competitors or enemies. Importantly, given de Waal's invocation of contemporary primatology against him, Huxley did in fact include empathy and a desire for fairness within the human natures that had then emerged from diverse evolutionary pressures. And chimpanzee empathy and fairness are the central capacities de Waal cites when describing morality's biological core. In fact, Huxley explicitly endorsed the biological precedents that (on de Waal's reckoning) veneer theorists are supposed to reject:

> I see no reason to doubt that, at its origin, human society was as much a product of organic necessity as that of the bees. The human family, to begin with, rested upon exactly the same conditions as those which gave rise to similar associations among animals lower in the scale. Further, it is easy to see that every increase in the duration of the family ties, with the resulting co-operation of a larger and larger number of descendants

for protection and defence, would give the families in which such modifications took place a distinct advantage over the others. And, as in the hive, the progressive limitation of the struggle for existence between the members of the family would involve increasing efficiency as regards outside competition.

<div align="right">(Huxley 1895: 26)</div>

This passage conclusively refutes de Waal's claim that Huxley ignored group selection, equated human existence with the struggle of each individual against the rest, and was therein led to equate moral speech with a disguised attempt to coerce essentially self-interested agents into cooperative endeavors. Huxley rejected that conception of morality whole cloth when he claimed, "that, at its origin, human society was as much a product of organic necessity as that of the bees."

Did Huxley claim the English were competing for survival and reproduction and so subject to the force of individual selection? Did he label his neighbors' calls for kindness, justice and patriotic solidarity a ploy used by them to gain advantage in their competition for greater progeny? Again the answer is "no":

> I think it would puzzle Mr. Lilly, or any one else, to adduce convincing evidence that, at any period of the world's history, there was a more widespread sense of social duty, or a greater sense of justice, or of the obligation of mutual help, than in this England of ours. Ah! but, says Mr. Lilly, these are all products of our Christian inheritance; when Christian dogmas vanish virtue will disappear too, and the ancestral ape and tiger will have full play. But there are a good many people who think it obvious that Christianity also inherited a good deal from Paganism and from Judaism; and that, if the Stoics and the Jews revoked their bequest, the moral property of Christianity would realise very little. And, if morality has survived the stripping off of several sets of clothes which have been found to fit badly, why should it not be able to get on very well in the light and handy garments which Science is ready to provide?

<div align="right">(Huxley 1895: 145)</div>

Of course, the call to war is often diagnosed as a hypocritical bid for the choicest mates, especially when it is loudly trumpeted by those who refuse to join in the fray. But one of Huxley's central conclusions in *Ethics and Evolution* was that natural selection had not affected the population of England over the course of the "four or five centuries" prior to his writing that work (1895: 40). In particular:

> During these three centuries, from the reign of Elizabeth to that of Victoria, the struggle for existence between man and man has been so largely restrained among the great mass of the population (except for one or two short intervals of civil war), that it can have had little, or no,

selective operation. As to anything comparable to direct selection, it has been practised on so small a scale that it may also be neglected. The criminal law, in so far as by putting to death, or by subjecting to long periods of imprisonment, those who infringe its provisions, prevents the propagation of hereditary criminal tendencies; and the poor-law, in so far as it separates married couples, whose destitution arises from hereditary defects of character, are doubtless selective agents operating in favour of the non-criminal and the more effective members of society. But the proportion of the population which they influence is very small; and, generally, the hereditary criminal and the hereditary pauper have propagated their kind before the law effects them. In a large proportion of cases, crime and pauperism have nothing to do with heredity; but are the consequence, partly, of circumstances and partly, the possession of qualities, which under different conditions of life, might have excited esteem and admiration.

(Huxley 1895: 39)

We are thus left with an interpretive mystery. Huxley clearly believed in group selection and the persistence of those biologically real forms of human sympathy, benevolence and justice for which it accounts. And Huxley clearly denied that individual selection was operating (or had operated) on English society to affect the distribution of psychological phenotypes therein. So Huxley explicitly posited a moral core and explicitly denied the then contemporary operation of those biological pressures that might favor the use of a "moral veneer" as a reproductive strategy. Why then does de Waal construe Huxley as the veneer theorist *par excellence*?

The answer, I think, can be found in Huxley's infamous rejection of the egalitarian ideology so movingly expressed by American and French revolutionaries in the latter half of the eighteenth century. Huxley is famous for arguing that this rhetoric was obviously false and bankrupt when interpreted in a descriptive sense: men are not "equal" and never have been (Huxley 1893a). Mightn't de Waal construe Huxley's critique of revolutionary ideology as evidence of Huxley's adherence to veneer theory? Of course, to reject egalitarian slogans is not yet to claim that morality's central function is coercive deception. Nor does it imply morality's biological or psychological irreality. But it's a start.

On this more nuanced reading of Huxley, "the veneer" he sees is not morality as a whole, but the expression of universal moral principles that are supposed to guide revolutionary political movements toward a truly egalitarian world order. What was pretense in Huxley's eyes was not a mother's "natural affection and sympathy" for her sick child or the Englishmen's quite genuine concern for the property rights of Englishmen wherever they may roam. These feelings are indeed both biologically and psychologically real (1895: 37).[10] But then so is our drive to dominate, our will to power:

The propounders of what are called the "ethics of evolution" when the 'evolution of ethics' would usually better express the object of their

speculations, adduce a number of more or less interesting facts and more or less sound arguments, in favour of the origin of the moral sentiments, in the same way as other natural phenomena, by a process of evolution. I have little doubt, for my own part, that they are on the right track; but as the immoral sentiments have no less been evolved, there is, so far, as much natural sanction for the one as the other. The thief and the murderer follow nature just as much as the philanthropist.

(Huxley 1895: 79–80)

What is mere pretense, according to Huxley, is not our evolved in-group morality, but the cosmopolitan idea that *we are all in the in-group*.

Though Huxley characterizes group selection as the main source of "the ethical progress" (1895: 35), he insists that ethical progress has a dark side: while it strengthens the loyalties, friendships and pro-social concerns of in-group members, it concurrently augments their hostility to members of various out-groups, who are in consequence conceptualized as "the enemies of the ethical":

Even should the whole human race be absorbed in one vast polity, within which "absolute political justice" reigns, the struggle for existence with the state of nature outside it, and the tendency to the return of the struggle within, in consequence of over-multiplication, will remain; and, unless men's inheritance from the ancestors who fought a good fight in the state of nature, their dose of original sin, is rooted out by some method at present unrevealed, at any rate to disbelievers in supernaturalism, every child born into the world will still bring with him the instinct of unlimited self-assertion. He will have to learn the lesson of self-restraint and renunciation. But the practice of self-restraint and renunciation is not happiness, though it may be something much better.

(Huxley 1895: 43–44)

Natural selection has not yielded wholly self-interested individuals who mask their wholly selfish motives in the course of their daily interactions. But the kind of biologically and psychologically real in-group solidarity that emerges from group selection is *inevitably* tied to out-group hostility: the kind of "othering" described by contemporary social theorists. As Hume so thoroughly argued, people tend not toward love of humanity itself but toward love of those humans with which they identify, however flexible the identities in question:

An Englishman in Italy is a friend: A Europoean in China; and perhaps a man wou'd be belov'd as such, were we to meet him in the moon. But this proceeds only from the relation to ourselves; which in these cases gathers force by being confined to a few persons.

(T 3.2.1.12, in Hume 2000)

But the wail of the chorus is real: *why* must we hate others to love one another? The answer, Huxley opines, is that we are naturally competitive in our acquisitiveness. We stop competing for resources with our brothers and sisters to help our family dominate the neighbors. We stop competing with the neighbors to help our village beat the next. And when all Englishmen have what they need to survive and reproduce without fear, when selective pressures fail to operate within the nation, "the struggle for existence, as between man and man, within that society, is ipso facto, at an end" (Huxley 1895: 36). But the competitive drives of the English were not *eliminated* during this period of social harmony. They were *displaced* or transformed into those that drove British imperialism. Might the kind of "othering" that accompanied this transformation provide the context in which Huxley found moral language functioning as a veneer?

I cannot address this interpretive question with the kind of scholarly care it should be given by a historian with deep knowledge of the period in question. But there are some suggestive facts worth reporting. As a first step, we might reconsider the passage I initially quoted from Huxley's *Evolution and Ethics* above, as it provides some insight into the mechanisms of self-deception that Europeans and settlers employed when laying a veneer of universal human rights over their engagement in racialized slavery. If you look back at that passage, you will see that Huxley there characterizes Native American tribes as components of the state of nature. He does not write of the natives as a rival population of people with their own set of norms and laws. He does not write of battles between European colonists and native populations as a clash of civilizations with Darwinian forces selecting the most "fit" group/ethos for survival and reproduction. Instead, Huxley describes colonization as a battle between civilization and *wilderness*, a battle he likens to a gardener's attempts to keep the weeds from choking off the plants of greater beauty and utility he has selected for himself. The natives are conceptualized as weeds rather than domestic animals; the results of natural rather than artificial selection. And this habit of thought wasn't unique to Huxley (Mills 1997).

Second, recall that John Locke (1988) used the idea that all men are by nature equal to argue against the authority of a Catholic monarch who lacked the consent of the English people he claimed to govern, and that Thomas Jefferson's justification for declaring independence from the British in 1776 invoked Locke's ideology when arguing for the self-evidence of a similarly universal set of principles: that all men are endowed by their creator with equal rights to life, liberty and the pursuit of happiness. We know too that Locke profited from the slave trade and was instrumental in writing inegalitarian (viciously inhumane) provisions into the slave code of the Carolinas (Bernasconi & Mann 2005), and that Jefferson "lived large" off his many slaves and failed to free them in his will (Cohen 1969; Berlin 1998). So Locke and Jefferson enslaved men while arguing loudly for the natural injustice of slavery. It is fair to say, then, that the universalist moral

language employed by Locke and Jefferson were components of a discursive façade they employed to hide their real lives.

These were veneers for sure, but not evidence for veneer theory as de Waal presents it. For what Locke and Jefferson were hiding in words was not amorality or immorality, but a *partial morality* limited to the wealthy Anglo-American clan of which they were both members. It is reasonable to suppose that Huxley knew of these instances of historically grand hypocrisy—and many others besides—and that this knowledge led him to conceptualize the revolutionaries' cosmopolitan ideology of universal human rights as pretense. It would also account for the viciousness of his critique of Rousseau.[11]

3. Conclusion

In his Tanner Lectures, de Waal indulges in a fairly gross form of interpretive unkindness. Huxley was not a veneer theorist in the sense de Waal there defines and then demolishes with the help of contemporary primatology. But despite his misreading, de Waal was nevertheless onto something. Huxley's (1893a) objections to the revolutionary rhetoric of natural equality can indeed be read as an endorsement of veneer theory. It's just that the theory in question utilizes the idea of a veneer in a much more subtle and defensible manner than de Waal pretends.

Notes

1 "One school views morality as a cultural innovation achieved by our species alone. This school does not see moral tendencies as part and parcel of human nature. Our ancestors, it claims, became moral by choice. The second school, in contrast, views morality as a direct outgrowth of the social instincts that we share with other animals. In the latter view, morality is neither unique to us nor a conscious decision taken at a specific point in time: it is the product of social evolution. The first standpoint assumes that deep down we are not truly moral. It views morality as a cultural overlay, a thin veneer hiding an otherwise selfish and brutish nature. Until recently, this was the dominant approach to morality within evolutionary biology as well as among science writers popularizing this field. I will use the term 'Veneer Theory' to denote these ideas, tracing their origin to Thomas Henry Huxley (although they obviously go back much further in Western philosophy and religion, all the way to the concept of original sin)" (de Waal 2003: 6–7).

2 De Waal also cites S. Freud (1913), the biologist G. C. Williams (1988) and the evolutionary theorist R. Dawkins (1976; 2003) as veneer theorists, and includes several social contract theorists discussed in the text below.

3 These issues are addressed in detail by D. Machuca's essay on J. L. Mackie's famous attempt to reconcile an "error theory" of moral discourse with a substantive first-order ethic (see Chapter 11, this volume).

4 When used to denote the extinction of one human group by another, "winnowing" has an objectionably euphemistic feel to it: *mutatis mutandis* for the "assimilation" of one culture into another.

5 On the role of early discourse in the evolution of morality, see Boehm (2014) and Tomasello (2014). Cf. Kitcher (2016: 192): "At some point between the

beginnings of ethical life and the invention of writing, our ancestors developed conceptions of self-regarding virtues, of social solidarity, and of respect for the law. It is, nevertheless, quite impossible to pinpoint the changes that occurred or to make responsible estimates of when they happened."

6 Cf. de Waal's (2003: 4–6) criticism of Rawls.

7 Huxley says that conceiving of the social contract as ideal rather than real was wise on Rousseau's part because Rousseau employs a "vicious method of a priori political speculation" which bears little connection to scientific biology and anthropology (Huxley 1893b: 336).

8 Huxley even countenanced species selection. "Man, the animal, in fact, has worked his way to the headship of the sentient world, and has become the superb animal which he is, in virtue of his success in the struggle for existence. The conditions having been of a certain order, man's organization has adjusted itself to them better than that of his competitors in the cosmic strife" (1895: 51).

9 It may be that Huxley joined Darwin in thinking that the colonists were more cooperative than the natives. And Huxley may have thought this advantage central to an explanation of colonialisms "success," i.e., the drastic evolution of phenotypes in the New World. According to Darwin: "A tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to aid one another, and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection. 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" (1982: 157–158; cf. 162–165). The texts with which I am familiar don't address Huxley's stance on these issues.

10 In fact, Huxley (1895: 31) endorsed Adam Smith's "impartial spectator" analysis of the sentiments that constitute an in-group's morality and the more neutral, abstract judgments of propriety, virtue and vice to which these sentiments give rise; and de Waal (2003: 12) classifies Smith as a paradigmatic *opponent* of veneer theory.

11 It is worth noting that the Anglo-American clan was not sufficiently united to sustain their trans-Atlantic romance. The American-British violently resented being looked down upon by the British-British. See Breen (1997) for discussion. Charitably interpreted, Mills' (1997) thesis is compatible with these observations: the racial contract was more pronounced in its effects (colonialism and slavery) than was the British contract in its (the revolution).

References

Berlin, I. 1998. *Many Thousands Gone: The First Two Centuries of Slavery in North America*. Cambridge, MA: Harvard University Press.

Bernasconi, R., & Mann, A. 2005. "The Contradictions of Racism: Locke, Slavery and the Two Treatises." In A. Valls (ed.), *Race and Racism in Modern Philosophy*, 89–107. Ithaca, NY: Cornell University Press.

Boehm, C. 2014. "The Moral Consequences of Social Selection." In F. de Waal, P. Churchland, T. Pievani, & S. Parmigiani (eds.), *Evolved Morality: The Biology and Philosophy of Human Conscience*, 31–48. Leiden: Brill.

Breen, T. H. 1997. "Ideology and Nationalism on the Eve of the American Revolution: Revisions Once More in Need of Revising," *The Journal of American History* 84: 13–39.

Byrne, R., & Whitten, A. 1988. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*. Oxford: Oxford University Press.

Cohen, W. 1969. "Thomas Jefferson and the Problem of Slavery," *Journal of American History* 56: 503–526.

Darwin, C. 1982 [1871]. *The Descent of Man, and Selection in Relation to Sex*. Princeton: Princeton University Press.

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

———. 2003. *A Devil's Chaplain: Reflections on Hope, Lies, Science, and Love*. New York: Houghton Mifflin.

De Waal, F. 2003. "Morality and the Social Instincts: Continuity with the Other Primates," *The Tanner Lectures on Human Values*, delivered at Princeton University.

———. 2006. *Primates and Philosophers: How Morality Evolved*. Edited by S. Macedo & J. Ober. Princeton, NJ: Princeton University Press.

Dugatkin, L. A. 1997. *Cooperation among Animals: An Evolutionary Perspective*. New York: Oxford University Press.

Freud, S. 1913. *Totem and Taboo*. New York: Norton.

Hume, D. 2000 [1738–1740]. *A Treatise of Human Nature*. Edited by D. F. Norton & M. Norton. Oxford: Oxford University Press.

Huxley, T. H. 1893a [1890a]. "On the Natural Inequality of Men." In Huxley 1893c, 290–335.

———. 1893b [1890b]. "Natural Rights and Political Rights." In Huxley 1893c, 336–382.

———. 1893c. *Collected Essays. Volume 1: Results and Methods*. London: Macmillan & Co.

———. 1895. *Evolution and Ethics and Other Essays*. London: Macmillan & Co.

Kitcher, P. 2016. "Evolution and Ethical Life." In D. L. Smith (ed.), *How Biology Shapes Philosophy: New Foundations for Naturalism*, 184–203. Cambridge: Cambridge University Press.

Lincoln, A. 1855. "Letter to Joshua Speed," www.abrahamlincolnonline.org/lincoln/speeches/speed.htm.

Locke, J. 1988 [1689]. *Two Treatises of Government*, 3rd edition. Edited by P. Laslett. Cambridge: Cambridge University Press.

Mills, C. 1997. *The Racial Contract*. Ithaca, NY: Cornell University Press.

Smith, D. L. 2011. *Less Than Human: Why We Demean, Enslave, and Exterminate Others*. New York: St. Martins Press.

Sterelny, K. 2012. *The Evolved Apprentice*. Cambridge, MA: MIT Press.

Tomasello, M. 2014. *A Natural History of Human Thinking*. Cambridge, MA: Harvard University Press.

Williams, G. 1988. "Reply to Comments on 'Huxley's Evolution and Ethics in Sociobiological Perspective'," *Zygon* 23: 437–438.

11 Moral Skepticism, Fictionalism, and Insulation

Diego E. Machuca

1. Introduction

In one of his now classic articles on skepticism, Myles Burnyeat (1997 [1984]) claims that a key difference between ancient and contemporary skeptics is that for the latter, ordinary beliefs and philosophical doubt are insulated from each other: one's ordinary beliefs about a given issue are unaffected by one's skepticism about that issue, and *vice versa*.[1] Given the theme of the present volume, the focus of this essay will be the relation between moral skepticism and first-order moral beliefs. Specifically, I will examine in what way first-order moral beliefs could be deemed to be insulated from second-order views that deny, or recommend suspension of judgment about, their truth or their epistemic justification. I am thus interested in only one of the directions in which insulation supposedly works.

The topic of moral insulation is related to a distinction commonly drawn between moral error theorists in relation to the attitude they should adopt towards ordinary moral thought and discourse, namely, that between *moral abolitionists* (or *eliminativists*) and *moral fictionalists*. Moral abolitionists maintain that moral error theorists should do away with first-order moral thought and discourse altogether: they should stop thinking according to moral categories and using moral language when talking with those who believe in the objectivity of morality. The reason typically adduced for the abolitionist recommendation is pragmatic: moralizing generates more personal and social costs than benefits. But one could also wonder whether the abolition of first-order moral thought and discourse is not logically or epistemically required by the eliminativists' moral anti-realism.[2] According to moral fictionalists, by contrast, given that morality produces many practical benefits (e.g., personal happiness or political stability), moral error theorists will be better off if they continue to make first-order moral utterances and have first-order moral thoughts, while at the same time refraining from asserting such utterances and believing such thoughts. When immersed in the moral fiction, the fictionalist is diverted from his moral skepticism and therefore acts as if he really believes in morality. But if pressed in a reflexive or critical context, he will recognize that it is nothing more than a fiction.

This incomplete description of moral fictionalism corresponds to the *revolutionary* kind, which is the fictionalist stance most commonly discussed in the literature and the one that is relevant to the topic of moral insulation. Whenever I speak of moral fictionalism, I will be specifically referring to revolutionary fictionalism. Although I will also take into account other alternatives available to the moral skeptic, my primary focus will be on fictionalism both because it has been endorsed by some of the most prominent contemporary moral skeptics, and because, in recommending that first-order moral thought and discourse be preserved, the moral fictionalist may be taken to believe that some kind of insulation is possible.

The aim of this essay is threefold. First, to assess whether the view that first-order moral beliefs are insulated from moral skepticism is defensible. Second, to examine whether a moral skeptic who adopts moral fictionalism can coherently hold that his first-order moral beliefs are unaffected by his skepticism about their truth or their epistemic justification. And third, to determine whether contemporary moral skeptics are in general committed to there being insulation between first- and second-order views.

I will begin by analyzing the notion of insulation: after presenting Burnyeat's treatment of it, I will distinguish between three main types of insulation (Section 2). I will then examine J. L. Mackie's stance on the phenomenon of insulation and his pragmatic conception of morality. In so doing, I will assess the plausibility of two of the three main types of insulation (Section 3). The reason for focusing on Mackie is not only that he is probably the best known and most important contemporary moral skeptic, but above all because in his *Ethics* one detects an illuminating tension between insulation and continuity between levels. Next, I will present moral fictionalism in more detail and consider whether any of the kinds of insulation distinguished in Section 2 is compatible with the adoption of a fictionalist stance. I will also compare moral fictionalism with a position known as "moral conservationism" (Section 4). I will close by summarizing the results of the examination of the above issues (Section 5).

Before getting down to business, I should remark that I will be concerned primarily with the error-theoretic form of moral skepticism—which claims that all first-order moral judgments are either false or neither true nor false because the objective moral facts or properties they purport to describe do not exist. The reason is that it is mainly moral error theorists who have claimed that first-order moral beliefs are insulated from moral skepticism, and who have embraced moral fictionalism. Still, I will also consider epistemological varieties of moral skepticism, namely, nihilistic epistemological skepticism—which asserts either that moral beliefs are epistemically unjustified or that moral knowledge is impossible—and Pyrrhonian skepticism—which consists in suspending judgment about whether moral beliefs are epistemically justified and about whether moral knowledge is possible.

2.   The Notion of Insulation

Burnyeat opens his essay on the phenomenon of insulation thus:

> Nowadays, if a philosopher finds he cannot answer the philosophical question 'What is time?' or 'Is time real?', he applies for a research grant to work on the problem during next year's sabbatical. He does not suppose that the arrival of next year is actually in doubt. Alternatively, he may agree that any puzzlement about the nature of time, or any argument for doubting the reality of time, is in fact a puzzlement about, or an argument for doubting, the truth of the proposition that next year's sabbatical will come, but contend that this is of course a strictly theoretical or philosophical worry, not a worry that needs to be reckoned with in the ordinary business of life. Either way he *insulates* his ordinary first order judgements from the effects of his philosophizing.
>
> (1997: 92)

Hardly anyone will deny that the phenomenon of insulation as depicted in this passage is widespread in present-day philosophy. This may be nothing more than the result of the fact that today philosophy is not normally conceived of as consisting in theoretical discussions leading to the discovery of the correct way of life, as used to be the case in antiquity. Be that as it may, the phenomenon of insulation is found not only among those academics engaged exclusively in the history of the discipline or the exegesis of philosophical texts—not to mention those who are mere 'doxographers'—but also among those engaged in so-called systematic philosophy. Take contemporary epistemological discussions of skepticism, which are in general characterized by a purely methodological use of skeptical arguments. It may be argued that, given that the great majority of contemporary epistemologists believe that skeptical arguments are unsound, the question of whether their ordinary beliefs are insulated from the conclusions of those arguments does not arise. Nevertheless, they recognize both the apparent plausibility of skeptical arguments (they are valid and their premises are intuitively compelling) and the difficulty in determining where those arguments go wrong—in fact, there is considerable disagreement about where exactly they go wrong. This recognition, however, has no effects whatsoever on their ordinary beliefs, which is seemingly due to the phenomenon of insulation. It is interesting to note that, if our first-order beliefs could not be affected by the conclusions of sound skeptical arguments, then pragmatic responses to skepticism would not make much sense, for in such a case skepticism would not really represent a threat to the attainment of those goals we take to be crucial to our lives.

   As regards moral skepticism, it is worth noting that, whereas few philosophers have in actual fact doubted or denied the reality of time or the existence of the external world or the possibility of having knowledge of

the world outside us, quite a few have doubted or denied the existence of objective moral facts or the epistemic justification of moral beliefs or the possibility of having knowledge of the moral realm. Moral insulation thus has the significance that, if it were a common phenomenon, it would present us with a stronger and clearer case than those found in other areas inasmuch as there would be quite a number of real skeptics whose first-order moral beliefs would not be infected by the second-order stances they adopt.

Returning to Burnyeat, he distinguishes two kinds of insulation: (*i*) insulation by subject matter or content (1997: 98–101, 110, 123), and (*ii*) insulation by level (1997: 122–123). Both types of insulation are discussed in connection with the debate about the scope of Sextus Empiricus's Pyrrhonism, namely, whether suspension of judgment is restricted to theoretical beliefs or extends also to ordinary or commonsense beliefs. In the case of insulation of type (*i*), commonsense beliefs are insulated from skepticism in that the latter targets a subject matter or a range of propositions with which ordinary people are not concerned in daily life. Insulation of type (*ii*) occurs when there are two distinct ways of understanding the very same proposition, namely, the ordinary way and the philosophical way. The skeptic does not target the plain man's knowledge claims, which are acceptable in the context of daily life, but only the philosophical claims to absolute knowledge of things as they are in themselves. Burnyeat does not refer to the relation between the two kinds of insulation, but it is clear that insulation by level is incompatible with insulation by subject matter in that, in the former, the philosopher is talking about the very same things as the ordinary man, but claims to possess a deeper understanding of them. Although the philosopher does not theorize about issues or problems that have nothing to do with the affairs of everyday life, ordinary beliefs about certain matters are unaffected by his theoretical views about those same matters. Insulation by content is uninteresting inasmuch as, if first-order beliefs are about matters different from those to which second-order views refer, then it is no mystery why the former are unaffected by the latter. I will therefore focus on insulation by level.

I think that the question of insulation by level should be approached by distinguishing between three main kinds of insulation: logical, epistemic, and psychological. I will describe each in relation to moral insulation. Logical insulation occurs if, from the truth or falsity of moral skepticism, one cannot logically infer the truth or falsity of first-order moral beliefs. Epistemic insulation occurs if knowing or justifiably believing that there are no moral facts or that there is no moral knowledge or no epistemically justified moral belief, or suspending judgment about such issues, does not imply that one should stop holding beliefs about the objective rightness or wrongness of certain actions. Psychological insulation occurs if, even though there is no logical or epistemic insulation, we are as a matter of fact hardwired in such a way that, even if we adopt moral skepticism, either we *cannot help* holding first-order moral beliefs, or we *may*

continue to hold such beliefs in case doing so turns out to be beneficial. Even though moral skepticism does *in principle* or *in theory* affect our first-order moral beliefs, *in practice* or *in fact* those beliefs either cannot but remain untouched or may become untouched by moral skepticism on account of our psychological makeup—perhaps the process of evolution has designed the human brain to engage in moral judgment.[3] Given this disjunctive characterization of psychological insulation, I distinguish between 'extreme psychological insulation' and 'mild psychological insulation'. Whereas extreme psychological insulation would be something that happens to us, mild psychological insulation would be the result of a decision we make for pragmatic reasons. One could perhaps further discriminate between two kinds of mild psychological insulation: one's first-order moral beliefs may become unaffected by moral skepticism *so long as* one keeps performing a certain mental action, or they may become unaffected by moral skepticism *once and for all* as soon as one performs such an action. Let me finally note that I will use the term 'insulator' to designate someone who thinks that his skepticism cannot affect his first-order beliefs, or who decides to shield his first-order beliefs from his skepticism for pragmatic reasons.[4]

3. Is Mackie an Insulator?

If you have perused Mackie's *Ethics*, you probably got the impression that something odd is going on: in the first part of a book someone proposes and defends a skeptical view according to which all substantive moral judgments are false because the objective moral facts they purport to describe do not exist, only to make in the second part claims that appear to express substantive moral judgments. The impression of oddness can be mitigated if Mackie thinks that there is no continuity between levels, and hence that moral skepticism—and metaethical views in general—cannot affect our first-order moral judgments. In fact, at times he explicitly embraces such a position, as when he distinguishes between first- and second-order moral skepticism. With regard to the former, he points out:

> '[M]oral scepticism' might also be used as a name for either of two first order views, or perhaps for an incoherent mixture of the two. A moral sceptic might be the sort of person who says 'All this talk of morality is tripe,' who rejects morality and will take no notice of it. Such a person may be literally rejecting all moral judgements; he is more likely to be making moral judgements of his own, expressing a positive moral condemnation of all that conventionally passes for morality; or he may be confusing these two logically incompatible views, and saying that he rejects all morality, while he is in fact rejecting only a particular morality that is current in the society in which he has grown up.
>
> (Mackie 1977: 16)

—

So Mackie distinguishes between three first-order skeptical views: (*i*) the rejection of all morality, (*ii*) the rejection of all conventional morality on the basis of a substantive moral judgment, and (*iii*) the incoherent rejection, on moral grounds, of all morality. The first of these is the most interesting for the topic of insulation, but before saying something about this, it must be observed that Mackie regards first- and second-order skeptical views as entirely separate from each other:

> These first and second order views are not merely distinct but completely independent: one could be a second order moral sceptic without being a first order one, or again the other way round. A man could hold strong moral views, and indeed ones whose content was thoroughly conventional, while believing that they were simply attitudes and policies with regard to conduct that he and other people held. Conversely, a man could reject all established morality while believing it to be an objective truth that it was evil or corrupt.
>
> (1977: 16)

It is plain that someone can coherently assert that morality as conventionally conceived is to be rejected while holding the second-order view that there are objective moral truths. The reason is that his first-order skeptical assertion is based on the substantive moral judgment that such morality is evil or corrupt (first-order moral skepticism of type (*ii*) above), and that this judgment in turn rests on that second-order view. Now, whereas there is no incoherence in this first case, how can someone retain strong moral views after denying that there are objective moral truths? In other words, how can someone keep holding beliefs about facts and properties that he claims do not exist, that are not part of, as Mackie likes to say, "the fabric of the world" (1977: 15, 22–24)? If first-order moral skepticism of type (*ii*) is independent of second-order moral skepticism because the former is based on a substantive moral judgment that in turn rests on a metaethical realist stance, how is it possible that substantive moral judgments like that one are not undermined by the adoption of a second-order skepticism that is the very denial of that realist stance?

Despite presenting a threefold taxonomy of first-order moral skepticisms in the first quoted passage, in the second Mackie only considers first-order moral skepticism of type (*ii*). Regarding first-order moral skepticism of type (*iii*), it is interesting that he rightly regards as incoherent the view that rejects, on moral grounds, all substantive moral judgments, but does not regard in the same way the view that maintains that such judgments are insulated from second-order skepticism. The reason is clearly that the first case concerns a relation between two first-order views, whereas the second case concerns a relation between a first- and a second-order view, which he regards as completely independent of each other. What about first-order moral skepticism of type (*i*)? From Mackie's formulation of it,

such skepticism corresponds to what is commonly called "moral nihilism," which is the first-order view that nothing is morally good or bad, right or wrong. It is considered a first-order view because it is expressed in the object language, which in the present case is the ordinary language used to make moral claims about 'objects' that exist in the world (actions, practices, events, persons, etc.). As I said above, this first-order moral skepticism is the most interesting one for the topic of insulation. The reason is that it might be argued that someone who rejects all morality or all substantive moral judgments as nonsense can legitimately do so only on the basis of a second-order moral skepticism, and hence that there is no insulation between his first- and second-order views. This seems to be precisely what distinguishes first-order moral skepticism of type (*i*) from that of type (*ii*): the latter type is based on a substantive moral judgment that in turn rests on a second-order realist stance, whereas the former type is based on a second-order moral skepticism that calls into question the objective truth of all substantive moral judgments. However, it is possible for someone to deny that anything is morally right or wrong without basing his view on some metaethical argument or without being able to defend his view when challenged by an opponent. Hence, endorsement of moral nihilism does not necessarily imply endorsement of second-order skepticism because the latter view is not necessarily endorsed on the basis of the former. This does not mean, however, that the reverse is true, for it is highly implausible that an error theorist, despite his considered ontological moral skepticism, is entitled to claim not to be a moral nihilist, i.e., entitled to continue to hold the belief that certain acts are morally right or wrong. The same goes for epistemological moral skepticism. Take Pyrrhonian moral skepticism, for example. One could suspend judgment about whether anything is morally right or wrong on the basis of no metaethical argument or without being able to defend his suspension if challenged. But it is highly implausible that someone who suspends judgment about the epistemic justification of both first-order moral beliefs and ethical theories is entitled not to suspend judgment about whether anything is morally right or wrong.

Mackie, then, takes first-order moral beliefs to be unaffected by second-order moral skepticism. Given the threefold distinction of types of moral insulation proposed in Section 2, what kind of insulator is he? Mackie explicitly opts for logical insulation in the following passage:

[R. M. Hare] sums up his case thus: 'Think of one world into whose fabric values are objectively built; and think of another in which those values have been annihilated. And remember that in both worlds the people in them go on being concerned about the same things—there is no difference in the "subjective" concern which people have for things, only in their "objective" value. Now I ask, "What is the difference between the states of affairs in these two worlds?" Can any answer be given except "None whatever"?'

Proof

> Now it is quite true that it is logically possible that the subjective concern, the activity of valuing or of thinking things wrong, should go on in just the same way whether there are objective values or not. But to say this is only to reiterate that there is a logical distinction between first and second order ethics: first order judgements are not necessarily affected by the truth or falsity of a second order view. But it does not follow, and it is not true, that there is no difference whatever between these two worlds. In the one there is something that backs up and validates some of the subjective concern which people have for things, in the other there is not.
>
> (1977: 21–22)

According to Mackie, then, even though in the moral skeptical world first-order moral beliefs are not epistemically justified because there are no objective values that "back up" and "validate" them, the moral skeptic can nonetheless retain those beliefs. For Mackie takes the subjective concern people have for things to consist not only in valuing things—which could be interpreted as merely expressing personal preferences—but also in thinking they are right or wrong and in making first-order moral judgments. And he explicitly maintains that there is a logical distinction between the two levels: our first-order moral judgments are not necessarily affected by the truth or falsity of moral skepticism, which is to be interpreted in the sense that their truth or falsity may be impervious to that of moral skepticism. Note also that, in making that logical distinction, Mackie is reiterating the point made when distinguishing earlier between first- and second-order skepticism, which means that thinking and judging that certain things are morally right or wrong consists in *holding strong moral views*. In sum, for Mackie there is a crucial ontological difference between the two worlds that has epistemic consequences in that, in the moral skeptic's world, the first-order moral claims are not backed up or validated by objective moral values. But such consequences do not extend so far as to prevent us from continuing to believe that things are morally right or wrong and to make first-order moral judgments expressing those beliefs. Given that first-order moral beliefs and judgments can be preserved, the logical insulation defended by Mackie has epistemic effects, and hence entails an epistemic insulation: knowing or justifiably believing that there are no mind-independent moral values does not imply that one should stop holding beliefs about the objective rightness or wrongness of certain actions.[5]

Another contemporary moral skeptic who explicitly endorses the view that first-order moral beliefs are insulated from second-order skepticism is Walter Sinnott-Armstrong—who, unlike Mackie, is not an error theorist, but an epistemological skeptic. For he maintains that "second-order beliefs about the epistemic status of moral beliefs cannot force us to give up the moral beliefs that we need to live well" (2006: viii). Moreover, he claims that moral skeptics

can hold substantive moral beliefs just as strongly as non-skeptics. Their substantive moral beliefs can be common and plausible ones. Moral skeptics can even believe that their moral beliefs are true by virtue of corresponding to an independent moral reality. All that moral skeptics deny is that their or anyone's moral beliefs are justified or known. . . . This meta-ethical position about the epistemic status of moral beliefs need not trickle down and infect anyone's substantive moral beliefs or actions.

(2006: 13–14)

Likewise, he points out: "I am *not* a moral nihilist. I believe that many acts are morally wrong. I think that my positive moral beliefs are true and correspond to moral facts" (2006: 58). My reason for quoting these passages is that Sinnott-Armstrong is even clearer and bolder than Mackie in his endorsement of the view that first-order moral beliefs are logically and epistemically immune from second-order moral skepticism. And like Mackie, Sinnott-Armstrong takes it as obvious that there is such insulation between levels—so obvious that it does not need to be argued for.

Notwithstanding this alleged obviousness, I confess that I fail to see the plausibility of logical and epistemic insulation. How is it logically and epistemically possible for someone to deny the "backup" or "validation" of a certain type of judgment while at the same time being entitled to make judgments of that type and to regard them as true? Likewise, how can someone, despite denying that his substantive moral beliefs are epistemically justified or known, hold to them as being true by virtue of their correspondence with objective or mind-independent moral facts? Consider the adoption of an error theory regarding witchcraft or astrology. It seems plain that, from the truth of an error theory about the existence of witches or the influence of the positions and movements of celestial bodies on earthly occurrences and human affairs, one can infer the falsity of a judgment that ascribes magical powers to a given person or of the predictions of today's horoscope. Similarly, one does not seem to be entitled to continue to hold first-order beliefs about witches or astrological matters if one claims to know that there are no witches or that the positions and movements of celestial bodies have no influence on earthly occurrences and human affairs. We can appreciate how implausible Mackie and Sinnott-Armstrong's view is if we think in terms of defeaters. For it can be argued that S's second-order skeptical view that his first-order moral beliefs are all false or epistemically unjustified serves as a defeater for S's first-order moral beliefs, and hence that S should stop holding them. Suppose that one forms one's beliefs about the temperature in a room at different times by looking at a thermometer on one of the walls. If one comes to believe that the thermometer has always been unreliable, then it seems plain that one's beliefs about the temperature are defeated and that one should therefore stop holding them. Likewise, suppose that one thinks one forms one's first-order moral beliefs via intuition. If one comes to believe that moral intuition is unreliable, then it seems plain that one's first-order

moral beliefs are defeated and that one should therefore stop holding them. If Mackie and Sinnott-Armstrong agreed that in the thermometer case the beliefs about the temperature are defeated and that one should therefore stop holding them, it seems they should agree that in the moral intuition case first-order moral beliefs are likewise defeated and that one should therefore stop holding them. And if they thought that first-order beliefs are defeated and should not therefore be held anymore in the former case but not in the latter, they would bear the burden of explaining why the two cases are different. It might be the case that moral skeptical arguments do not in fact have an impact on our first-order moral beliefs due to some form of psychological insulation, but this is of course different from claiming that they should not, which is what proponents of logical and epistemic forms of insulation, such as Mackie and Sinnott-Armstrong, maintain.

In the passages of *Ethics* quoted thus far, Mackie explicitly endorses the view that first-order moral beliefs are logically and epistemically insulated from second-order moral skepticism. There are, however, several passages in which he seems to accept that the moral skeptic's second-order stance does have an effect on the first-order level, thus infecting his ordinary moral beliefs. Consider the following:

> 'Our sense of justice,' whether it is just yours and mine, or that of some much larger group, has no authority over those who dissent from its recommendations or even over us if we are inclined to change our minds. But if there is no objective moral truth to be discovered, is there nothing left to do but to describe our sense of justice?
>
>   At least we can look at the matter in another way. Morality is not to be discovered but to be made: we have to decide what moral views to adopt, what moral stands to take. No doubt the conclusions we reach will reflect and reveal our sense of justice, our moral consciousness. . . . But that is not the object of the exercise: the object is rather to decide what to do, what to support and what to condemn, what principles of conduct to accept and foster as guiding or controlling our own choices and perhaps those of other people as well.
>
> (1977: 105–106)

> [T]he content of the first order moral system is more malleable, more a matter of choice, than utilitarianism, in any form, makes it appear . . . there is no merit in pretending that our choices are rationally constrained in ways that they are not. We are, then, free to mould or remould our moral system so as better to promote whatever it is that we do value.
>
> (1977: 146)

> It does not follow . . . that an individual is free to invent a moral system at will. If a morality is to perform the sort of function described

in Chapter 5, it must be adopted socially by a group of people in their dealings with one another.

<div align="right">(1977: 147)</div>

To say that someone has a right, of whatever sort, is to speak either of or within some legal or moral system: our rejection of objective values carries with it the denial that there are any self-subsistent rights.

<div align="right">(1977: 173)</div>

Morality as I have described it is concerned particularly with the well-being of active, intelligent, participants in a partly competitive life, and the constraints summed up as morality in the narrow sense have been introduced . . . as necessary limits on competition for the benefit of all the competitors.

<div align="right">(1977: 193)</div>

[T]he arguments of the preceding chapters . . . show how there can be a secular morality, not indeed as a system of objective values or prescriptions, but rather as something to be made and maintained, and which there is some real point in making.

<div align="right">(1977: 227)</div>

The rationality of morality . . . consists in the fact . . . that men need moral rules and principles and dispositions if they are to live together and flourish in communities, and that evolution and social tradition have given them a fairly strong tendency to think in the required ways.

<div align="right">(1977: 228–229)</div>

The idea that morality must be made, invented, or molded rather than discovered is central to Mackie's view, as already indicated in the very subtitle of his 1977 book: *Inventing Right and Wrong*. What is the status of this morality? Insofar as the error theorist is aware that his own morality has no authority over those who have adopted a different one—given that there is no objective prescriptivity on the basis of which he is entitled to claim that they are required to endorse his own morality—it seems that the first-order moral beliefs that he used to hold before becoming a moral skeptic are indeed affected by his second-order stance. This is why saying that someone has a right makes sense only within a given invented (legal or moral) system: the denial of the existence of objective moral values entails the denial of the existence of self-subsistent rights. The adoption of moral error theory does not therefore leave things as they were: morality as ordinarily understood is undermined or debunked, and so we must make and maintain a new one. If first-order moral beliefs remained intact in the face of moral skepticism, what need would there be for constructing a new first-order moral system? There are, according to Mackie, restrictions on which first-order moral

system to invent and adopt, but these are pragmatic, not moral, restrictions: the first-order moral system must make it possible to live within a society in a way that benefits all participants; it must provide rules that regulate behavior and govern social interactions in a way that allows everyone to flourish.[6] One may quite reasonably wonder in what sense we can continue to call such a system a 'moral' one insofar as it merely reflects our preferences, those things that we in fact value, and not objective values, prescriptions, or requirements. Besides the fact that the moral skeptic continues to experience what can be described as moral emotions, the answer seems to lie mainly in the fact that the first-order moral systems of the skeptic and the realist perform a certain function: the commonality is not found in the ontologico-epistemological foundation of their moral systems, but on the personal and social aims they help to achieve. In addition, the skeptic's first-order moral system reflects, just as the realist's, our strong tendency, shaped by evolution and culture, to think according to moral categories. Hence, Mackie's conception of such a first-order moral system is a pragmatic one, which again shows that the undermining or debunking effects of his moral error theory do carry over to the first-order level, infecting ordinary moral beliefs. This exclusively pragmatic conception of morality would dispel the impression of oddness to which I referred at the very beginning of the present section: the second part of *Ethics* does not propose a first-order moral system that is of the same kind as those targeted by moral error theory. Such a pragmatic conception of morality is incompatible with moral insulation and squares well with moral fictionalism—a stance to be discussed in the next section. In fact, at one point Mackie does talk of morality being a useful fiction: In so far as the objectification of moral values and obligations is not only a natural but also a useful fiction, it might be thought dangerous, and in any case unnecessary, to expose it as a fiction. This is disputable. (1977: 239).

In a later book, while recognizing the significant benefits of morality insofar as it fulfills a social function, Mackie (1980: 154–156) also emphasizes its negative effects, but without going as far as to recommend the adoption of moral abolitionism. It is possible that, in the end, he remained in a state of suspension of judgment about whether morality should be abolished or retained as a fiction.

Before concluding my discussion of Mackie, I would like to mention a peculiar interpretation of his stance on insulation that is suggested by Enoch (2011: 42–43): though Mackie's moral error theory has first-order implications, it does not have *discriminating* first-order implications. This means that moral error theory establishes that morality is a fiction, but has no implications within that fiction: it does not allow one to make distinctions between claims made within the moral fiction or to settle disputes that arise within it.[7] This interpretation will not do because it accepts that moral error theory affects all first-order moral judgments inasmuch as these judgments can be made only within the context of the moral fiction. But

the logical and epistemic insulation that Mackie advocates at certain points (and that Sinnott-Armstrong wholeheartedly embraces) requires more than that: it requires that, despite one's moral skepticism, one can continue to hold beliefs about the objective rightness or wrongness, and not merely the within-the-fiction rightness or wrongness, of certain actions. If ordinary moral discourse loses its claim to moral objectivity, then it is not immune from moral skepticism.

Mackie seems at times to accept that there is continuity between first- and second-order levels—which would explain his pragmatic conception of morality—thereby creating a strong tension with the logical and epistemic insulation he explicitly claims to embrace at other times. If there is indeed such a tension in Mackie, I confess that I do not know how to resolve it or if it can be resolved. But leaving this interpretive issue aside, one of the points I want to make in this essay is that moral insulation is incompatible with certain views commonly adopted by contemporary moral skeptics, and hence that it would be a mistake to affirm that all or most of them are insulators.

4.   Moral Fictionalism and Insulation

In this section, I will discuss moral fictionalism and one of its competitors. My aim is not to assess their plausibility or to decide between them, but to examine their compatibility with the view that first-order moral beliefs are insulated from moral skepticism. My description of moral fictionalism will be based on the version defended by Richard Joyce, because it is both the best known and the one relevant to the topic at hand.[8] Remember that whenever I speak of moral fictionalism, I have in mind the revolutionary kind.

What does moral fictionalism consist in? The first thing to say is that the moral fictionalist is a skeptic. To the best of my knowledge, in the metaethical literature moral fictionalism has always been associated with a moral error-theoretic position, but it could in principle be adopted by any moral skeptic—except the moral non-cognitivist, since the moral fictionalist takes the default use of moral language to be assertoric or descriptive. So the moral fictionalist is someone who could in principle be either an ontological or an epistemological moral skeptic: he denies that there are objective moral facts or properties, or denies that we have moral knowledge or justified moral beliefs, or suspends judgment about whether there are objective moral facts or properties and about whether there is such a thing as moral knowledge or justified moral beliefs. In what follows, I will focus for the most part on the moral fictionalist who is an error theorist.

The second aspect of the moral fictionalist's stance to be mentioned concerns his practical attitude towards first-order moral thought and discourse once morality has been debunked—the attitude that defines him as a *fictionalist*. He maintains that it would be irrational to carry on holding moral beliefs, that it is not psychologically possible to do so, and that, even if it

were psychologically possible, one should not carry on holding them because truth is of instrumental value. Nevertheless, the moral fictionalist does not claim that we should do away with morality altogether, because he believes that morality is useful inasmuch as, by bolstering self-control, it allows us to attain certain personal and social goals. He therefore proposes to preserve first-order moral thought and discourse in a way that does not commit us to error (or lying or self-deception), namely, by having moral thoughts without believing them and by making moral utterances without asserting them, in the same way in which a storyteller thinks about and utters propositions he knows would be false were he to use them to describe, predict, or explain what he takes to be reality. The moral fictionalist maintains that even when moral thoughts are knowingly employed as fictions, they can nevertheless engage our emotions—in the same way in which reading a story or seeing a movie can engage them—and that emotional reactions can influence our motivations to act in ways conducive to attaining desirable ends. So in his daily life the moral fictionalist carries on employing first-order moral discourse as if it were not false, immersing himself in the moral fiction and pretending that it is true. But if pressed in a critical or reflective context—either by others or by himself—he will immediately step out of the fiction and recognize it as such. It is not that in the ordinary context he stops endorsing his moral skepticism; it is just that he is not attending to it. The moral fictionalist does not affirm that his stance provides all the benefits of a believed morality, but only that, after examining the costs and benefits of his stance and its alternatives,[9] it is moral fictionalism that seems to get the better results. There is thus a pragmatic reason for the moral fictionalist's decision not to dispense with first-order moral thought and language altogether, and so in a purely instrumental sense, it can be said that he is justified in maintaining the fiction that moral realism is true.[10]

This brief characterization of moral fictionalism makes it clear that the reason why its proponent continues to use first-order moral language is not that he believes that his moral skepticism does not affect first-order moral judgments. Rather, precisely because moral skepticism does have a defeating effect on first-order moral judgments, the moral fictionalist refrains from using moral language in an assertoric or descriptive way. He reasons and acts *as though* moral judgments were true and *as though* he held first-order moral beliefs, without losing sight of the fact that this is what he is doing. Hence, moral fictionalism is incompatible with both logical and epistemic forms of insulation. Similarly, if, as a matter of psychological fact, one's first-order moral beliefs cannot but be insulated from one's moral skepticism or if one could insulate them from it in case it turned out that holding such beliefs is beneficial, then there would be no need to pretend to believe in the fiction that certain actions, practices, or events are morally right or wrong in an objective sense. Moral fictionalism is thus also incompatible with both extreme and mild psychological moral insulation. It could be argued, though, that there is a sense in which the moral

fictionalist accepts moral insulation: he recognizes that some aspects of morality—namely, moral thoughts and moral emotions—are for the most part immune from moral skepticism. Even though I think this is correct, it should be noted that such a kind of insulation does not leave things entirely as they were: first-order moral beliefs and assertions are done away with on account of the skeptical arguments that undermine their epistemic credentials. The kind of insulation that is intriguing, and on which I focus in this essay, is *doxastic* insulation, i.e., that of first-order moral beliefs from second-order skepticism. Moral fictionalism is incompatible with any kind of doxastic insulation. Hence, if a moral error theorist adopted moral fictionalism, he could not claim, at the risk of being inconsistent, that his first-order moral beliefs are insulated from his moral skepticism, for there could not be such separation between the two levels. The tension in Mackie's position to which I called attention at the end of the previous section can now be construed more clearly as a tension between moral insulation and moral fictionalism. Note also that the apparent incongruity between the two parts of *Ethics* mentioned at the outset of the previous section can be explained either by Mackie's endorsement of logical and epistemic insulation or by his endorsement of moral fictionalism.

To shed more light on moral insulation and its incompatibility with moral fictionalism, one can compare the latter with moral conservationism, a position recently endorsed and defended by Jonas Olson (2014: ch. 9). He proposes it as an alternative available to the moral error theorist that is better than either moral abolitionism or moral fictionalism. Like the moral fictionalist and unlike the moral abolitionist, the moral conservationist thinks that morality is socially useful in that it allows us to prevent and resolve conflicts, regulate interpersonal relations, and counteract limited sympathies. What distinguishes moral conservationism? Olson describes it as the "preservation of ordinary (faulty) moral thought and discourse" (2014: 178). But thus formulated, the view does not clearly differ from moral fictionalism. For the moral fictionalist, too, recommends that we keep employing ordinary moral thought and discourse, albeit not in the same way as they are normally used inasmuch as the ordinary man does believe and assert the moral sentences he utters. We must therefore interpret moral conservationism as the *complete* preservation of ordinary moral thought and discourse, both regarding the content of moral sentences and the illocutionary force with which they are uttered. Hence, the moral conservationist maintains that we should keep believing and asserting first-order moral sentences even though we know full well that they are all false. One can interpret the moral conservationist as endorsing a restricted form of irrationality: one should keep holding beliefs one knows to be false due to their instrumental value—a paradigmatic case of self-deception.

Olson describes the moral conservationist's attitude as a sort of compartmentalization: "conservationism recommends moral belief in morally engaged and everyday contexts and reserves attendance to the belief that

moral error theory is true to detached and critical contexts, such as the philosophy seminar room" (2014: 192). Olson maintains that this compartmentalization is feasible:

> [I]t is a psychologically familiar fact that we sometimes temporarily believe things we, in more reflective and detached contexts, are disposed to disbelieve. In such cases, the more reflective beliefs are suppressed or not attended to. This might be because of emotional engagement, affection, peer pressure, or a combination of these factors. [. . .]
> Something similar might be going on with moral beliefs. The error theorist might say, 'I knew all along there is no such thing as moral wrongness, but hearing on the news about the massacre on [*sic*] civilians, I really believed that what the perpetrators did was wrong; I really believed that the UN ought morally to enforce a cease fire'. [. . .]
>   [C]ertain actions and events may engage our emotions of anger, empathy, etc., to the effect that it seems to us that the actions are morally wrong and that we virtually cannot help believing that they are morally wrong, no matter how intellectually compelling we find arguments in favour of moral error theory. It appears realistic that in morally engaged and engaging contexts, affective attitudes like anger, admiration, empathy, and the like, tend to silence beliefs that moral error theory is true.
> (2014: 192–193)

And in a note he offers two analogies:

> Some optical illusions are such that it seems to us that one line is longer than another, even though we know that the lines are of equal length. It seems possible that in unreflective moments we believe, on the basis of how things seem, that one line is longer. [. . .] Many utilitarians who are convinced that their theory is correct and that according to this theory, the bystander ought to push the fat man off the bridge in the famous trolley case . . . still feel an intuitive reluctance to make this judgement.
> (2014: 193, n. 42)

The first thing to note is that the moral fictionalist who is an error theorist, a nihilistic epistemological skeptic, or a Pyrrhonist will not disagree with Olson that certain actions or events still trigger in them both moral thoughts (or 'appearances', as the Pyrrhonist would call them) and emotional reactions. For instance, a Pyrrhonian fictionalist does not express a belief when, in referring to a stick half-submerged in water, he remarks: "It appears to me that the stick is broken." Similarly, when he says "It appears to me that *x* is morally wrong," he is not expressing a belief about *x*, but only the way he is still affected by *x* on account of, e.g., his upbringing, education, and life experiences. Olson talks indistinctively, and erroneously, of moral beliefs and what we might call 'moral seemings' or 'moral appearances', which is precisely

what allows him to use the two analogies. Also, it is clear that the emotions brought about by certain actions or events cause the moral fictionalist to make utterances that express approval or disapproval, and to act accordingly, but there is no reason to assume that, while making such utterances, he believes that their contents describe objective moral facts or properties.

Leaving aside my reservations about Olson's position, what is important for present purposes is that, although he does not refer to insulation as such, it is plain that his moral conservationism is the explicit endorsement of the view that first-order moral beliefs can be shielded from moral error theory. Because holding first-order moral beliefs generates practical benefits, the moral conservationist makes the decision to insulate them from his skepticism. The kind of insulation in question is therefore what I called "mild psychological insulation" in Section 2: the moral conservationist does not claim that first-order moral beliefs are immune *from* moral error theory, but rather that they can become immune *to* it if one comes to the conclusion that preserving them is beneficial, and as long as one engages in a process of compartmentalization.[11] Why is the insulation in question not of the logico-epistemic kind? Because if our first-order moral beliefs were logically or epistemically immune from moral error theory, there would be no need to compartmentalize those beliefs and the belief in the truth of that theory. If in ordinary contexts the moral conservationist should not attend to the conclusions of the sound arguments in favor of moral error theory for pragmatic reasons, it is because such conclusions do have a defeating effect on his first-order moral beliefs.[12]

It could be argued that the apparent incongruity between the two parts of Mackie's *Ethics* is to be explained by the fact that he endorses moral conservationism. That Mackie is a moral conservationist is suggested by Olson himself, according to whom Mackie could be taken

> to be saying that a man can subscribe to both views [that moral discourse is error-ridden and that we can hold strong moral beliefs], but that he cannot attend to both views simultaneously. He must compartmentalize his thoughts and avoid entertaining both thoughts at the same time.
>
> (2014: 41–42, n. 75)

The problem with this interpretation of Mackie's position is simply that, in the passages in which he talks about the lack of continuity between levels, the form of insulation in question is logico-epistemic, not psychological. Even though, like Olson, Mackie thinks that morality is useful in that it serves a social function, pragmatic reasons have nothing to do with his claim that first-order moral beliefs are unaffected by second-order moral skepticism.

Someone might argue that fictionalism actually rests upon a kind of psychological insulation, in a way similar to moral conservationism. For when

the moral fictionalist is in an ordinary context using moral terms and thinking in accordance with moral concepts, he is in fact insulating his moral discourse and thinking from his moral skepticism. However, even though there seems indeed to be some kind of psychological *dissociation* at work here, it is clearly not a type of *insulation*, for the moral fictionalist does not adopt the *schizophrenic doxastic attitude* of the moral conservationist. There is dissociation inasmuch as the moral fictionalist is not *attending to* his moral skepticism while immersed in the moral fiction, for this might interfere with his fictive attitude. If, while watching a science fiction movie, I keep telling myself that what I am watching is incredible or highly implausible or absurd, doing so might interfere with my enjoyment of the story being told. But there is no insulation in the moral fictionalist's stance inasmuch as, while immersed in the moral fiction, he does not *hold* moral beliefs or *assert* moral utterances because he does *not abandon* his moral skepticism. Not only does he recognize that the moral fiction is a fiction if asked in the context of a philosophical discussion, but he also does so recognize if asked out of the blue in the context of ordinary life. If, while watching a science fiction movie, I am asked whether I believe that most of the things being depicted are possible, I will immediately reply that of course I do not believe so. Once again, the very formulation of a moral fictionalist stance presupposes that the conclusions of certain metaethical skeptical arguments do affect our first-order moral claims and prevent us from continuing to believe those claims.

5.   Concluding Remarks

Let me sum up the results obtained in the previous sections. First, in analyzing the notion of insulation, it might be appropriate to distinguish between logical, epistemic, and psychological insulation. Regarding the first two forms of insulation, it is not defensible to claim that first-order beliefs are logically or epistemically insulated from second-order views, because first-order beliefs can be defeated by second-order views on the basis of higher-order evidence. Psychological insulation, by contrast, seems more plausible inasmuch as, if it occurred, it would be a fact about our psychological makeup: as a matter of psychological fact, certain first-order beliefs either are always impervious to second-order views or may become so in certain circumstances.

Secondly, there is in Mackie, probably the most prominent contemporary moral skeptic, a strong tension between the view that first-order moral beliefs are logically and epistemically insulated from second-order moral skepticism, on the one hand, and the recognition of continuity between levels that leads him to a pragmatic conception of morality, on the other. This tension can be interpreted as a tension between moral insulation and moral fictionalism, which are clearly incompatible. Indeed, given that moral fictionalism maintains that the moral skeptic should proceed *as though* he believed that

first-order moral claims are true, moral fictionalists admit that, from the truth of moral skepticism, one can infer the falsity of first-order moral judgments, and that one should stop holding first-order moral beliefs once one becomes a moral skeptic. Consideration of the mild psychological insulation described by the moral conservationist makes such incompatibility clearer as far as this form of insulation is concerned. Note also that examination of moral fictionalism and moral conservationism shows that extreme psychological insulation does not occur inasmuch as proponents of those views are able to abandon their first-order moral beliefs once they realize that these beliefs are defeated by their skepticism. By contrast, one can say that mild psychological insulation does occur if one believes the moral conservationist's report on his experience of self-deception. The immunity conferred by the process of compartmentalization is not attained once and for all, but will be retained as long as the moral conservationist does not attend to his moral skepticism in ordinary contexts. One could hypothesize that those moral skeptics who mistakenly think that their first-order moral beliefs are logically and epistemically immune from their moral skepticism think so at least in part because they experience mild psychological insulation: they are able to insulate their substantive moral beliefs from their skepticism when they recognize their pragmatic value—as we saw, even Sinnott-Armstrong talks about "the moral beliefs that we need to live well."

Finally, it would be a mistake to claim that contemporary moral skeptics are in general committed to the view that there is insulation between their first-order beliefs and their skepticism. The moral fictionalist is clearly not an insulator in any of the senses distinguished—Mackie being a complex and intriguing case—and the mild psychological insulation described by the moral conservationist is not a widespread phenomenon. The moral abolitionist's stance, too, is incompatible with moral insulation: if one can and should abolish morality, then it is not the case that first-order moral beliefs are logically, epistemically, or psychologically insulated from moral skepticism. Note also that moral skeptics who adopted propagandism (on which see notes 9 and 12) or what has recently been called "revolutionary expressivism" (Köhler & Ridge 2013) would not be insulators either. The propagandist would not try to keep the truth of moral skepticism from the general public if he thought that first-order moral beliefs are unaffected by it. And the revolutionary expressivist who claims that if moral error theory is true, we should become moral expressivists, accepts the moral error theorist's view that ordinary moral claims express beliefs but maintains that they should no longer express beliefs in case the antecedent of the conditional obtains. If first-order moral beliefs were impervious to moral skepticism, there would be no need to adopt an expressivist or non-cognitivist stance. That the view that there is insulation between first-order beliefs and skepticism is not prevalent among contemporary moral skeptics should probably come as no surprise, given that it is most likely the result of the fact that we are talking about real skeptics and not merely philosophers who examine

and discuss skepticism in a detached manner and do not in the end take it seriously. It seems that when skepticism becomes a real option, the view that our first-order beliefs are insulated from it dissipates considerably.[13]

Notes

1 Burnyeat's thesis is not restricted to a difference between ancient and contemporary skepticism, but refers to a difference between ancient and contemporary philosophy more generally. However, Burnyeat's whole treatment of insulation and all of his examples concern skepticism, and in any case moral skepticism is the topic in which I am interested.

2 To the best of my knowledge, all contemporary moral eliminativists are moral anti-realists: see Hinckfuss (1987), Garner (1994; 2010), Burgess (2010), and Marks (2013). But nothing seems to prevent an epistemological moral skeptic from being a moral eliminativist.

3 What I describe as logical insulation and epistemic insulation may be partially similar to what Bett (1993: 375–377) calls "extreme insulation," while what I describe as psychological insulation may be somewhat close to what he calls "practical insulation" (1993: 374–375, 377). Also, while the first two kinds of insulation I distinguish may be similar to what Wong (2002: 350) defines as the insulation that concerns "a certain relation between scepticism and common sense," the third kind of insulation I identify clearly corresponds to what he characterizes as the insulation that concerns "a certain relation between the sceptic's beliefs about the world and his sceptical belief" (2002: 350).

4 Let me make an idiomatic digression. There is a difference sometimes drawn between "immune from" and "immune to." If x is immune from y, then x is exempt from y or is not subject to y because x cannot be touched by y. By contrast, if x is immune to y, then x is resistant or impervious to y because, although y may touch x, y has no effect on x (see Garner 2003: 430). For example, whereas humans are immune from diseases that affect only birds, they can become immune to a given human disease if they get vaccinated either once or every few years. One may argue that logic, epistemic, and extreme psychological forms of insulation occur if first-order moral beliefs are immune *from* moral skepticism, whereas mild psychological insulation occurs if they are immune *to* it.

5 Both Burnyeat (1997: 112) and Bett (1993: 378, 380) maintain that Mackie (1977) is an insulator, although the former only quotes and examines a short passage in which Mackie does not actually endorse insulation, and the latter does not examine any of the relevant passages of Mackie's work.

6 Mackie (1982: 246–247, 251, 254) reiterates that "value itself is a human and social product," and that the "invention of moral values" has made it possible for us to better live together, survive, and flourish.

7 Probably motivated by the healthy caution of any good systematic philosopher, Enoch (2011: 43 n. 51) remarks that he does not "do history, recent history included," and so he offers his suggestion because it is of independent interest, not because he believes it is a good interpretation of Mackie. By saying that he does not do history, what he means in this case is of course that he does not do exegesis, not even of not-so-distant twentieth-century authors. I wonder how such exegesis differs from interpreting the views defended in the latest article or book of a present-day philosopher with whom one will never be in contact.

8 See Joyce (2001: ch. 8; 2016b: 58–66; 2016c). Cf. Nolan *et al.* (2005).

9   The alternatives are to do away with morality altogether (abolitionism or eliminativism), to both believe and promulgate belief in morality despite the evidence of its falsehood (conservationism, on which more below), and to hush up the evidence for moral error theory (what Joyce [2001: 214] calls "propagandism").

10 Joyce's moral fictionalism is of the revolutionary stripe because it recommends a radical change in one's attitude towards moral discourse once one becomes an error theorist. Mark Eli Kalderon (2005), in contrast, defends "hermeneutic" moral fictionalism, according to which ordinary people already adopt a make-believe or fictive attitude towards moral discourse. (Nolan *et al.* [2005] call these two varieties of fictionalism "prescriptive" and "descriptive," respectively.) Hermeneutic fictionalism is a form of non-cognitivism, and a peculiar one at that, since it claims that, although moral sentences do express propositions that attribute moral properties to things or that represent putative moral facts, the acceptance of a moral sentence is not belief in the moral proposition expressed and the utterance of a moral sentence is not the assertion of the moral proposition expressed. The question of insulation does not arise for hermeneutic fictionalists, nor for traditional non-cognitivists: given that they contend that first-order moral claims do not express beliefs, it would make no sense for them to either affirm or deny that first-order moral beliefs are unaffected by moral skepticism.

11 Although he himself is a moral non-cognitivist, in the course of discussing moral error theory, Kalderon (2005: 103–105), too, seems to accept psychological insulation. For he holds that endorsing moral error theory does not entail that we should abandon, or suspend judgment about, first-order moral beliefs: it might be rationally permissible to continue to hold such beliefs either because (*i*) it is psychologically impossible to stop holding them, or because (*ii*) it is socially beneficial to keep them. I interpret (*i*) as extreme psychological insulation and (*ii*) as mild psychological insulation (if there were pragmatic reasons for keeping first-order moral beliefs despite their falsity, it would be psychologically possible to do so).

12 Olson (2014: 196 n. 48) erroneously claims that his conservationism is similar to the view Cuneo and Christy (2011) call "propagandism"—actually, they call it "propagandism in the broad sense" because it is an expansive version of the view Joyce describes as propagandism (see note 9 above). For the propagandist in the broad sense takes up *non-doxastic* attitudes towards moral propositions (Cuneo & Christy 2011: 94–95, 101). Conservationism seems to correspond to what they call "intransigentism" (Cuneo & Christy 2011: 93). Olson's mistake is due to the fact that, just as the conservationist, the propagandist in the broad sense does not propose to transform ordinary moral discourse, but "to more or less leave things as they are" (Cuneo & Christy 2011: 101). But note that the fictionalist does not propose to transform ordinary moral discourse either, but only recommends that those who have become moral error theorists adopt a fictive attitude when using first-order moral language.

13 I am grateful to Dale Chock, Hallvard Hillehammer, Nate King, and Aaron Zimmerman for their helpful suggestions and critical remarks.

References

Bett, R. 1993. "Scepticism and Everyday Attitudes in Ancient and Modern Philosophy," *Metaphilosophy* 24: 363–381.

Burgess, J. 2010. "Against Ethics." In Joyce & Kirchin 2010, 1–15.

Burnyeat, M. 1997 [1984]. "The Sceptic in His Place and Time." In M. Burnyeat & M. Frede (eds.), *The Original Sceptics: A Controversy*, 92–126. Indianapolis: Hackett.

Cuneo, T., & Christy, S. 2011. "The Myth of Moral Fictionalism." In M. Brady (ed.), *New Waves in Metaethics*, 85–102. Basingstoke: Palgrave Macmillan.

Enoch, D. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.

Garner, B. 2003. *Garner's Modern American Usage*. Oxford: Oxford University Press.

Garner, R. 1994. *Beyond Morality*. Philadelphia: Temple University Press.

———. 2010. "Abolishing Morality." In Joyce & Kirchin 2010, 217–233.

Hinckfuss, I. 1987. *The Moral Society: Its Structure and Effects*. Camberra: Australian National University.

Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.

———. 2016a. *Essays in Moral Skepticism*. Oxford: Oxford University Press.

———. 2016b. "Morality, Schmorality." In Joyce 2016a, 41–66.

———. 2016c. "Moral Fictionalism." In Joyce 2016a, 219–239.

Joyce, R., & Kirchin, S. (eds.). 2010. *A World without Values: Essays on John Mackie's Moral Error Theory*. Dordrecht: Springer.

Kalderon, M. 2005. *Moral Fictionalism*. Oxford: Oxford University Press.

Köhler, S., & Ridge, M. 2013. "Revolutionary Expressivism," *Ratio* 26: 428–449.

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

———. 1980. *Hume's Moral Theory*. London & New York: Routledge.

———. 1982. *The Miracle of Theism*. Oxford: Oxford University Press.

Marks, J. 2013. *Ethics without Morals: A Defence of Amorality*. New York: Routeldge.

Nolan, D., Restall, G., & West, C. 2005. "Moral Fictionalism versus the Rest," *Australasian Journal of Philosophy* 83: 307–330.

Olson, J. 2014. *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press.

Sinnott-Armstrong, W. 2006. *Moral Skepticisms*. New York: Oxford University Press.

Wong, W. 2002. "The Problem of Insulation," *Philosophy* 77: 349–373.

Contributors

David Copp is Distinguished Professor of Philosophy at the University of California, Davis. He is the author of *Morality, Normativity, and Society* (Oxford University Press, 1995) and *Morality in a Natural World* (Cambridge University Press, 2007), and he has edited several anthologies, including *The Oxford Handbook of Ethical Theory* (Oxford University Press, 2006). He is editor of a monograph series with Oxford University Press called "Oxford Moral Theory." He has published and lectured widely on topics in moral and political philosophy.

Terry Horgan is Professor of Philosophy at the University of Arizona. He works in metaphysics, epistemology, philosophy of mind, and metaethics. He is the co-author of *Connectionism and the Philosophy of Psychology* (Bradford/MIT Press, 1996), *Austere Realism: Contextual Semantics Meets Minimal Ontology* (MIT Press, 2008), and *The Epistemological Spectrum: At the Interface of Cognitive Science and Conceptual Analysis* (Oxford University Press, 2011). He has also co-edited *Metaethics After Moore* (Oxford University Press, 2006).

Richard Joyce is Professor of Philosophy at Victoria University of Wellington. He works principally in metaethics and moral psychology. He is the author of *The Myth of Morality* (Cambridge University Press, 2001), *The Evolution of Morality* (MIT Press, 2006), and *Essays in Moral Skepticism* (Oxford University Press, 2016). He is editor of the *Routledge Handbook of Evolution and Philosophy* (Routledge, 2017) and has also co-edited *A World Without Values: Essays on John Mackie's Moral Error Theory* (Springer, 2010) and *Cooperation and Its Evolution* (MIT Press, 2013).

Hallvard Lillehammer is Professor of Philosophy in the School of Social Sciences, History and Philosophy at Birkbeck, University of London. He is the author of *Companions in Guilt: Arguments for Ethical Objectivity* (Palgrave, 2007), and the co-editor of *Real Metaphysics* (Routledge, 2003) and *Ramsey's Legacy* (Oxford University Press, 2005).

Diego E. Machuca is Associate Researcher in Philosophy at CONICET (Argentina) and co-editor of the *International Journal for the Study of*

Skepticism. His research focuses on skepticism in ancient philosophy, epistemology, and metaethics. He has edited *Pyrrhonism in Ancient, Modern, and Contemporary Philosophy* (Springer, 2011), *New Essays on Ancient Pyrrhonism* (Brill, 2011), and *Disagreement and Skepticism* (Routledge, 2013). He has also co-edited *Skepticism: From Antiquity to the Present* (Bloomsbury, 2018) and *Les raisons du doute: études sur le scepticisme antique* (Classiques Garnier, 2018).

Bart Streumer is Professor of Philosophy at the University of Groningen. His main research interest is metaethics. He is the author of *Unbelievable Errors* (Oxford University Press, 2017) and the editor of *Irrealism in Ethics* (Wiley-Blackwell, 2014). His articles have appeared in such journals as *Australasian Journal of Philosophy*, *Journal of Moral Philosophy*, *Philosophical Studies*, *Philosophy and Phenomenological Research*, and *The Journal of Philosophy*.

Folke Tersman is Professor of Philosophy at Uppsala University (Sweden). He works mainly in metaethics and is the author of *Moral Disagreement* (Cambridge University Press, 2007). His recent publications include "Disagreement: Ethics and Elsewhere" (*Erkenntnis* 79, 2014), "Explaining the Reliability of Moral Beliefs" (in U. Leibowitz and N. Sinclair (eds.), *Ethics and Explanation*, Oxford University Press, 2016), and "Debunking and Disagreement" (*Noûs*, forthcoming).

Christine Tiefensee is Associate Professor of Philosophy at the Frankfurt School of Finance & Management. She specializes in metaethics and broader questions of normativity. Her current research concerns inferentialist approaches to metaethics, relaxed moral realism, and the normativity of rationality. Recent publications include "Inferentialist Metaethics, Bifurcations and Ontological Commitment" (*Philosophical Studies* 173, 2016).

Mark Timmons is Professor of Philosophy at the University of Arizona and editor of *Oxford Studies in Normative Ethics*. His research interests include Kant's ethics, metaethics and moral phenomenology, and moral epistemology. He is the author of *Morality Without Foundations: A Defense of Ethical Contextualism* (Oxford University Press, 1998) and *Significance and System: Essays on Kant's Ethics* (Oxford University Press, 2017). He has also co-edited *Metaethics After Moore* (Oxford University Press, 2006).

Mark van Roojen is Professor of Philosophy at the University of Nebraska-Lincoln. His main work is in ethics, metaethics, and political philosophy, but his philosophical interests are wider yet. He is the author of *Metaethics: A Contemporary Introduction* (Routledge, 2015). His most widely read papers discuss noncognitivism, moral rationalism, moral psychology, normative epistemology, and normative semantics.

Aaron Zimmerman is Associate Professor of Philosophy at the University of California, Santa Barbara. His research is focused on the intersection between thought, language, and reason. He is the author of *Moral Epistemology* (Routledge, 2010) and *Belief: A Pragmatic Picture* (Oxford University Press, 2017). His articles have appeared in such journals as *Hume Studies*, *Philosophical Studies*, *Synthese*, and *The Philosophical Quarterly*.