No need to get up from the armchair (if you're interested in debunking arguments in metaethics)

Dan Baras[1] 📷

Abstract

Several authors believe that metaethicists ought to leave their comfortable armchairs and engage with serious empirical research. This paper provides partial support for the opposing view, that metaethics is rightly conducted from the armchair. It does so by focusing on debunking arguments against robust moral realism. Specifically, the article discusses arguments based on the possibility that if robust realism is correct, then our beliefs are most likely insensitive to the relevant truths. These arguments seem at first glance to be dependent on empirical research to learn what our moral beliefs are sensitive to. It is argued, however, that this is not so. The paper then examines two thought experiments that have been thought to demonstrate that debunking arguments might depend on empirical details and argues that the conclusion is not supported.

1 The question

There is an ongoing debate about the following claim:

> *Conditional Debunking Conclusion*: If robust moral realism is correct, then *none* of our moral beliefs are epistemically justified.

Robust moral realism is the view that there are irreducible non-natural objective moral truths (Enoch 2011, 1). As I explain in the sections that follow, robust moral realism in conjunction with a package of theses that typically go with it, is more prone to debunking challenges in comparison to other metaethical views, including other variants of realism. The consequent of Conditional Debunking Conclusion is a global undercutting of the justification of our moral

---

✉ Dan Baras
dan.baras@mail.huji.ac.il

[1] Martin Buber Society of Fellows, Mandel School for Advanced Studies in the Humanities, The Hebrew University of Jerusalem, Mt. Scopus, 9190501 Jerusalem, Israel

beliefs. If both robust moral realism and Conditional Debunking Conclusion are correct, that means we are hopeless about morality. We can't attain justified moral beliefs, not to mention moral knowledge which requires justification. There is a different debate about more local debunking of specific moral beliefs. Such local debunking is not the concern of this article.

Conditional Debunking Conclusion plays a central role in recent Metaethics. Michael Ruse (1986, chap. 6), Richard Joyce (2001, chap. 6, 2016) and Sharon Street (2006, 2015) argue for Conditional Debunking Conclusion and then use it to argue that robust moral realism is untenable because it has such a harsh implication. Other authors have followed their path, developing refined arguments.[1] Their arguments have prompted ongoing debate. The task of this paper is not to settle the debate. Rather, in this paper I wish to examine the following question regarding this debate: Should it be conducted from the armchair? Or, conversely, must we be significantly empirically informed in order to rationally assess the Conditional Debunking Conclusion? Within the expanding moral debunking literature, one can discern the emergence of two opposing schools of thought on this matter. The first school believes that philosophers must get their hands dirty with empirical details. The following quote is exemplary:

> [A] frustrating trend has emerged as [evolutionary debunking arguments] against morality have been more and more widely debated, namely, a tendency to abstract away from relevant empirical detail. (Fraser 2014, 458).[2]

A second school of thought believes that empirical details make little difference in these debates.[3] This article attempts to provide some support for the second school. I argue that a very superficial knowledge of science gives us all the information we need to conduct our inquiry from the armchair.

The plan for the remainder of this paper is as follows. In the next section I characterize robust moral realism and explain why, thus characterized, it is more prone to debunking challenges than competing metaethical views. In section 3 I discuss a debunking argument that, more than others, invites the thought that empirical details matter. I argue that they don't. Then, in sections 4 and 5, I examine thought experiments that are put forward by Derek Parfit and Joshua Greene as counterexamples to the thesis of this paper. These authors believe that their examples demonstrate that empirical details matter. I argue that their thought experiments don't in fact support the intended conclusion.

---

[1] Surveys of the debate include Vavova (2015) and Wielenberg (2016). This debate has spilled over to a number of additional domains, such as religion and metaphysics. For a comprehensive survey, see Korman (2019).

[2] Similar claims are made by Hopster (2018), Isserow (2019), Levy & Levy (2018), Mogensen (2016, 1812) and Parfit (2017, 3:286). In the context of a more limited debunking argument, Greene (2014, 726) expresses a similar attitude. Pölzler (2018), May (2018) and Sauer (2018) are examples of whole books devoted to engaging with detailed empirical science and debunking arguments in ethics of various sorts.

[3] Articles in this spirit include Chappell (2017), Kahane (2011, 111–12), Klenk (2017) and White (2010). Berker (2009) focuses on Greene's dual process theory and Greene and Singer's claim that it undermines deontological reasoning. I do not discuss these elements here. All of these authors' claims are of limited scope. They do not explicitly claim that empirical details never matter. Nevertheless, I suggest that they are examples of a school of thought or philosophical leaning that tends to disregard the importance of empirical details for debunking arguments in metaethics.

An interesting side note: Cuneo (2007) suggests that the division between these empirical vs. non-empirical inclinations underlies the division between metaethical naturalists and non-naturalists as well.

2 Unpacking the robust realist's package

In this section, I explain what I mean by robust moral realism and why I focus on this view. I do so by laying out a number of theses that are endorsed by robust realists. The first of these theses is essential to the view. The rest, while not logically entailed by the definition of robust realism, are theses that are typically endorsed by robust realists, and for good reasons, as I explain. I therefore consider them all as part of the moral realist's package.

The first idea is that morality, as a domain of truths, is in a meaningful and interesting way independent, or autonomous, from other domains. There are two autonomy theses that I have in mind here, one metaphysical and the other epistemic. The *metaphysical autonomy* thesis is the claim that the truth makers of moral claims and non-moral claims are very different.[4] What it takes for claims about how the world ought to be, or about how good or bad the world is, or about which actions are right or wrong to be true, is very different from what it takes for claims about how the world actually is, or about the non-normative properties of the world, or about which actions are or are not actually performed to be true. The metaphysical claim is closely related to the Moorean idea that the good is unanalyzable (Moore 1903, secs. 10–14), though the precise relationship between Moore's semantic claim and the metaphysical claim depends on some further theoretical questions that I do not delve into. A common way of stating the thesis is to claim that moral properties are *sui generis*. Metaphysical autonomy is an essential part of robust realism, since it is part of what robust realists mean when they say that moral facts are non-natural (Parfit 2011, 2:324; Enoch 2011, 100–109).[5]

A more recent and helpful way of stating the thesis is that moral facts are never fully grounded in non-moral facts (Rosen 2017). Any moral fact is either fundamental, i.e. ungrounded, or else at least partially grounded in a moral fact. This, according to Rosen, is what distinguishes non-naturalists from naturalists in metaethics. It follows that we can distinguish between pure moral facts and impure moral facts. *Pure* moral facts are either fundamental or are fully grounded in moral facts. *Impure* moral facts are partially grounded in contingent non-moral facts. (Moral facts that are grounded in a conjunction of moral facts and other necessary facts I count as pure moral facts).[6] There is some controversy as to whether pure moral facts are metaphysically necessary or contingent.[7] However, even according to the view that they are contingent, it is quite uncontroversial that changes in contingent non-moral facts cannot change any *pure* moral facts. If <murder of an innocent person is impermissible>

---

[4] David Enoch calls the intuition that motivates this view, the *just-too-different* intuition. For a clear and concise survey on how the *just-too-different* intuition challenges naturalism, see Paakkunainen (2018).

[5] Shafer-Landau (2003, 72) also endorses metaphysical autonomy as part of his non-naturalism. However, when he discusses the meaning of "non-naturalism" (p. 58) he prefers a disciplinary characterization rather than metaphysical. According to Shafer-Landau, non-naturalism means that morality does not belong to the same discipline as the natural sciences, primarily because the methodology for discovering moral truths is different from the methodology of natural science. Thanks to Preston Werner and to Russ Shafer-Landau himself for helping clarify Shafer-Landau's views.

[6] Moral facts are facts about how we ought to act or about which states of affairs are good or bad. Therefore, pure moral facts, even though purely moral, will still be about non-moral circumstances, that is, a conjunction of non-normative properties. For instance, the claim "it is bad to cause pain" tells us that *causing pain*, a non-normative circumstance, has the moral property *bad*.

[7] Kit Fine (2002) argues that there is a unique normative necessity. Fine's view is further developed by Gideon Rosen (forthcoming) who also nicely clarifies the relationship between this view and the widely accepted view that moral properties strongly supervene on non-normative properties. If all pure moral facts are metaphysically contingent, that means that moral properties do not *strongly* supervene on non-moral properties, though weak supervenience is not ruled out. (The distinction between weak and strong supervenience is made by Kim (1984)).

is a pure moral fact (I'm not saying that it is. I'm just using it as a hypothetical illustration), then no matter what our world might be like in its non-moral aspects, it would still be impermissible to murder an innocent person.

Metaphysical autonomy is typically a significant part of the motivation to accept *epistemic autonomy*, which is the Humean claim that non-moral information about a state of affairs is insufficient to support a belief in a substantial moral claim about that state of affairs.[8] The slogan version is that you cannot derive an ought from an is. When Fitzpatrick, for example, explains what he means by the autonomy of ethics, it becomes clear that he means the epistemic thesis:

> [E]thics is essentially autonomous. By this I mean that truths in ethics—truths about morality, reasons for acting, excellence of character, what is good for a person, and so on—can be accurately and justifiably arrived at only through engaged first-order ethical reflection and argument employing its own internal standards, and not from the outside through some other form of inquiry, such as psychology or biology. (FitzPatrick 2008, 172).[9]

The two autonomy theses described motivate two stronger theses that will figure in my argument. The same intuition that underlies metaphysical autonomy leads to the thought that:

> *Non-causality*: Moral facts cannot stand in causal relationships.

Of course, the mere fact that moral facts are different does not in itself entail that they are causally inert.[10] Rather, with regard to moral facts, the same intuition that motivates meta-physical autonomy, the idea that they are different, also suggests that they cannot stand in causal relations. Facts about what we ought to do just do not seem like the kind of thing that can cause anything.[11]

Epistemic autonomy leads to the thought that:

> *Intuitions are our only hope*: If we have any hope of justifying our moral beliefs, it is by relying on our moral intuitive judgments as evidence. We have no other source of moral input.

How? Epistemic autonomy implies that from any non-moral source of information, such as empirical information, you cannot infer any moral conclusions. There must be some source of moral input to epistemically ground moral beliefs. And moral intuitions seem like the only

---

[8] Finlay (2007, n. 25) also notes this link between the two types of autonomy. Hume himself seems to have based his epistemic thesis on the metaphysical intuition, that "*ought*, or *ought not*, expresses some new relation… entirely different from [*is*, and *is not* propositions]" (Hume 1888, sec. 3:1:1).

[9] See also Shafer-Landau (2003, 63).

[10] To see this clearly, consider: Mind and body are very different according to dualism, but they do causally influence each other. Perhaps dualism is an implausible view, however, it is not inconsistent.

[11] I am assuming that the term "non-naturalism" that I use to characterize robust realism does not logically imply non-causality. Shafer-Landau (2003, 58–59) and Cuneo (2007, 856) think similarly. That is why I have to explain why, even though it is not a logical necessity, non-naturalists typically do endorse non-causality.

Our *beliefs* about what we ought to do can of course affect our behavior. I am assuming that, according to robust realism, the belief-states are distinct from the facts themselves, such that this would not be considered a way of moral facts causing things.

things that fit the job.[12] So, if you accept epistemic autonomy, it is natural, though not necessary, to conclude that moral intuitions are our only hope.

Notice that accepting these claims does not yet imply that empirical information is irrelevant for debunking arguments. Even if you think that relying on our intuitions is our only hope, that does not mean that we are justified in relying on our intuitions. It may turn out that empirical research, coupled with certain epistemic norms, imply that we should not rely on our moral intuitions. That is precisely what the insensitivity argument is supposed to establish.

What happens if you reject the autonomy theses? Some ways of rejecting the autonomy theses imply that there is no real debunking worry. This is why robust realists are more prone than others to debunking arguments. If you reject metaphysical autonomy, that is, if you think moral facts are not disconnected in the relevant way from other facts, then you should likely reject epistemic autonomy. Once you reject epistemic autonomy (whether or not this is a consequence of rejecting metaphysical autonomy) then that means you do not think that some unique epistemic means is required to establish moral beliefs. Therefore, you will probably end up with a view that implies that you need not rely on suspect moral intuitions to discern moral truths and there is little worry about debunking. Let me demonstrate how this could work with a recent example. Sterelny and Fraser (2016) argue that moral truths are just truths about enhancing cooperation. (Even if this is a crude way of putting their view, I think the point still holds for the more fine-grained view). If that is the case, then we have other means of determining which actions are likely to enhance cooperation, such as modeling and experimenting.

More generally, much of the motivation to reject autonomy theses is a general naturalistic tendency, a reluctance to accept into our ontology anything that is not supported by empirical science. Therefore, people are tempted to somehow locate moral truth within the naturalistic ontology. Now, if moral truths are connected in this way to the empirical world, it will probably be the case that we can have just as much confidence about moral claims as we can about anything empirical. After all, the tighter connection with empirical beliefs is what motivates such views to begin with. To the extent that we are not worried about the debunking of our empirical beliefs, we should not be worried about our moral beliefs. I do not claim that this remark is true of all theories that reject the autonomy theses. I just want to flag that regarding such theories, it is worth examining whether it applies.

As I formulated these claims, they are particularly about moral truths. It might be the case that other normative truths are similar to moral truths in these respects. Perhaps, for instance, reasons for belief are just as autonomous as moral reasons for action. This is surely not the case for all norms. Norms of chess playing or of etiquette are dependent on conventions. They are not autonomous in either sense.[13] To keep things simpler, in this paper I focus merely on moral norms and remain uncommitted regarding other types of norms.

---

[12] Some people think that certain religious authorities or texts fit the job as well. I do not, for predictable reasons, but I do not get into this debate here. Whether moral perceptualists should be understood as rejecting this thesis is an interesting question that I do not pursue here. For a presentation of the view and related discussion, see Bergqvist and Cowan (2018) and Werner (2016, 2017).

[13] See Paakkunainen (2018, sec. 3) for a brief survey. Tristram McPherson (2011, 232) distinguishes between robust normativity and merely formal normativity. Formal normativity means that actions (etc.) can be right or wrong according to some standard, where the standard can be conventional, made up or anything else. Robust normativity requires that the standard be authoritative. As John Mackie (1977, 40) put it with regard to moral norms, they must have *to-be-doneness* built into them. It seems plausible to conjecture that the autonomy theses, if true, are true of all and only robust norms. Thanks to David Enoch for the reference.

My argument will be that if you accept these claims, as robust realists typically do, then empirical details cannot make a significant difference to the question at hand.

3 Arguments from insensitivity

How might empirical research confirm or disconfirm the Conditional Debunking Conclusion? If you accept the autonomy theses, then you shouldn't expect empirical science to be able to tell us how often our moral judgments are or are likely to be correct. Empirical research can't do so without some independent way of determining what the moral truths are. According to robust realists of the kind we are examining, our own moral judgments are our only hope at determining what the moral truths are. Trivially, all of our moral judgments are in line with our moral judgments. In this section I examine a more promising route.[14]

One influential type of debunking argument is premised on an epistemic principle that when we receive information that a given belief is insensitive to the relevant truth, or, using a different term, does not track the relevant truth, that belief thereby loses its epistemic justification. Indeed, the most influential evolutionary debunking arguments are put in terms of sensitivity or tracking.[15] An important feature of this line of thought is that it seems empirical research may be able to determine what our moral beliefs are sensitive to without any assumptions about what the moral truths are. If we learn that our moral beliefs are sensitive to factors other than the moral truths, that would seem to debunk their justification. Here is the general form of the argument:

(1)   (I learn that) the processes that produce my moral beliefs are insensitive to the relevant truths.
(2)   If (I learn that) the processes that produce a set of beliefs are insensitive to the relevant truths, then those beliefs are not epistemically justified.
(3)   Therefore, my moral beliefs are not epistemically justified.

Why do I add the "I learn that" clause in parentheses? It seems plausible that a belief can be justified even if it is insensitive to the relevant truth, so long as the subject is justifiably not aware of this fact.[16] For example, if I look at a clock, and the clock says 12:06, I am justified in

---

[14] There is a different style of argument developed by Hartry Field (1989, 25–30) as an argument against mathematical Platonism, and applied more recently to robust moral realism (Enoch 2011, chap. 7; Street 2008). According to this argument, robust moral realism should be rejected because it implies that the reliability of our moral beliefs, which is a fact that calls for explanation, cannot be explained. This argument is not premised on the Conditional Debunking Conclusion. I engage with Field's style of argument in depth elsewhere and provide reasons to think that it too does not depend on empirical details. I therefore allow myself to set it aside here. I engage with Field's style of argument in depth elsewhere and provide reasons to think that it too does not depend on empirical details (Baras 2017).

[15] See Ruse (1986, 254); Joyce (2001, 163); Sinnott-Armstrong (2006, 43); Street (2006); Bedke (2014); Joyce (2016); Hill (2016); Braddock (2017); May (2018, chap. 4) and Sauer (2018, chap. 2). For an influential response, see Clarke-Doane (2016, sec. 2.2). Clarke-Doane does so on the basis of a principle which he calls "modal security". For discussion of that principle, see Clarke-Doane & Baras (2019).

[16] As you can tell from this claim, my inclinations are internalistic. Others might be more attracted to a kind of externalism according to which what matters is whether my belief is sensitive, not whether I know it to be so. According to this view, the parenthetical additions should be omitted.
   What happens if you learn that your beliefs are insensitive, but you also justifiably believe that insensitivity and even knowledge of insensitivity don't undercut justification? I believe this is an interesting question, one I cannot explore it here.

believing that the time is 12:06 even if the clock is broken, so long as I do not know that the clock is broken and could not have known that the clock is broken. However, once I learn that the clock is broken, I am no longer justified in believing that the time is 12:06 on the basis of looking at the broken clock, even if, by lucky coincidence, the time is in fact 12:06. Learning that the clock is broken constitutes an undermining defeater for my belief about the time.[17]

As mentioned earlier, I focus on the sensitivity argument because it seems to me most prone to invite the thought that the empirical details matter. It is tempting to believe that detailed empirical information is needed to discern whether the first premise is correct.[18] Moral psychologists sometimes describe some of their results as discovering that our moral judgments are or are not sensitive to this or that factor. It seems therefore that we must turn to empirical research to discover what our moral judgments are sensitive to.[19] In this section, I argue that such empirical data has no epistemic impact on sensitivity-based debunking arguments. I do not determine whether the argument is sound. Rather, I point out a number of possible pathways one might go in determining whether premise (1) is true. Some of these pathways imply that it is true, others that it is not. My point will be that either way, empirical details are not going to make a difference.

What does it take for a belief forming process to be sensitive to a relevant truth? The term "sensitive" is ambiguous. One possibility is that sensitivity implies a causal relationship. For example, my sensory beliefs about the shape and color of my desk are the products of sensitive processes. Why? Because of a causal relationship of the right type between my desk and my brain.[20] Rays of light bounce off my desk in a certain way, and travel to my retina. They then have an effect on photoreceptors that are *sensitive* to light. Those photoreceptors send signals to neurons and somehow that translates into a reliable visual experience. There is a relevant causal chain. We can define insensitivity accordingly:

> *Appropriate causal connection*: The belief that p is insensitive to the relevant truth iff there is no appropriate causal connection between p and the belief that p.

Some people have doubts that information about insensitivity in the causal sense undermines beliefs. One problem is that mathematical beliefs as well as beliefs about the future seem justified, despite the fact that in these cases, the content of the belief (p) is not among the causes of the belief (Bp).[21] However, we can revise the appropriate causal connection principle

---

[17] Does that mean that the beliefs of most of us won't be undermined in this way, because we lack detailed knowledge of the mechanisms and genealogy of the processes that produce our moral beliefs? Not at all. I remind you that, I'm about to argue that all that is needed here is a very minimal acquaintance with science, the kind that the vast majority of us, definitely the readers of this article, do have.

[18] I do not imply that the second premise does not need examination as well, only that empirical details do not even initially seem relevant to this examination. For discussion of versions of the second premise, see White (2010, 580–82).

[19] Note that the target of this paper is independent from and sometimes confused with a different project. We have quick gut feelings and it is a great empirical project to find out when and why our quick judgments match up with our slower more thoughtful judgments. I think this is a charitable interpretation of Woodward and Allman (2007, 2008), Railton (2014) and some of Greene's (2013, 2014) work. The former two defend a more optimistic view of intuitions; Greene gives a more mixed verdict. However, the kind of debunking that is the target of this paper is the kind that undermines our moral judgments as a whole, including our slow and more thoughtful judgments.

[20] The qualification "of the right type" is needed to exclude deviant causal chains.

[21] Benacerraf's original argument against mathematical Platonism was based on this kind of causal condition, applied to knowledge. The causal condition has since significantly fallen out of favor, which is part of why philosophers of mathematics tend to focus more on Field's development of the argument, which does not rely on a causal condition on knowledge or justification. Regarding Field's style of argument, see above footnote 14.

to accommodate such beliefs.[22] And, at least to some, the revision does not seem too ad hoc. Mathematical truths are truths that all causal mechanisms must obey, including the mechanisms responsible for our mathematical beliefs. So, there is a causal link between mechanisms that obey mathematical laws and our mathematical beliefs, even if it would be awkward to say that the mathematical laws in themselves cause anything. In addition, mathematics is an essential part of our scientific explanations. Mathematical claims therefore will have to be part of the ultimate causal explanation of our beliefs, even if they are not causes in themselves. Future events are not part of the cause of our beliefs about the future. Rather, we make inferences about the future from our knowledge about the laws or statistical regularities in nature and the current state of affairs. These in turn are part of the causes of the future events. Hence, with regard to beliefs about the future, the same facts that determine future events are part of the causes of our beliefs about the future. Therefore, even if the relevant facts do not cause the beliefs in these cases, there is a kind of causal relationship between them.

The same cannot be said for moral truths, if you accept the non-causality thesis of the previous section. Moral facts play no causal role whatsoever, and it seems they cannot be part of any genuine causal explanation either.[23] Thus, if sensitivity implies a causal relationship, we know that moral beliefs cannot be sensitive to moral facts. No additional empirical information is needed to establish this conclusion.

David Faraci ([2019](#)) has recently suggested that a causal requirement on justification or knowledge is too stringent, and we should be epistemically satisfied with any appropriate explanatory connection between the belief and the relevant truth.[24] That is, the explanatory connection need not be causal, if, for example, a truth is metaphysically grounded at least in part in the belief, that should be enough. An interpretation of the sensitivity requirement along Faraci's lines would be as follows:

> *Appropriate explanatory connection*: The belief that p is insensitive to the relevant truth iff there is no appropriate explanatory connection between p and the belief that p.

However, we can rule out, from the armchair, the possibility of maximal scientific information having an impact here as well. We already ruled out the possibility of a causal connection between our moral beliefs and robust moral truths. We can also rule out the possibility of a metaphysical grounding relationship between moral beliefs and the robust moral truths. Robust realism, as defined above (with Rosen's help), entails that moral beliefs do not even in part ground moral truths. Nor does it seem plausible that the moral truths metaphysically ground, even in part, moral beliefs. For one, it is not part of what it is to be a moral belief that it conforms to a moral truth. More importantly, non-causal explanatory relationships are just not part of what scientists discover. (Perhaps they do discover mathematical explanations, but surely not moral explanations).

---

[22] For a recent development of a debunking argument along these lines, see Lutz ([forthcoming](#)).

[23] Sturgeon ([1984](#), [2006](#)) famously argues in favor of moral explanations. However, Sturgeon is not a robust realist; hence his view lies outside the scope of this paper. For a survey, and doubts about the claim that moral non-naturalists should reject moral explanations, see Majors ([2007](#)).

[24] Faraci's main concern is with epistemic coincidence. However, in a footnote (n. 32) he suggests that there is a tight connection between epistemic coincidence and epistemic justification. In particular, that evidence of epistemic coincidence is undermining. This implies that a lack of explanatory connection between belief and truth is undermining, as *Appropriate explanatory connection* says. This idea, in connection to debunking arguments, has recently been further explored by Korman & Locke ([forthcoming](#)).

While the non-causality of moral facts implies that we cannot be sensitive to pure moral facts in the first sense of sensitive, it does not imply that we cannot be sensitive to some relevant truths for impure moral facts. Recall that impure moral facts are partially grounded in contingent non-moral facts. Now we can be causally sensitive to such non-moral facts such as whether another person is in pain or whether somebody is lying or telling the truth. You may think that empirical data can inform us as to how sensitive we are to non-normative facts that partially ground impure moral facts. Perhaps, for instance, we can learn that we are not as good as we initially thought in detecting when certain people are in pain.

However, such empirical information would surely be far from establishing the Conditional Debunking Conclusion we are concerned with here. First, because as bad as we may turn out to be, it seems quite obvious (from the armchair) that we are not doomed when it comes to being sensitive to the relevant non-normative properties, such as other people's pain.[25] Second, because, as already noted, such information would not have any implications with regard to pure moral beliefs. And since impure moral facts are always partially grounded in more fundamental pure moral facts, if the scientific information cannot help us with regard to the latter, we have not made much progress.

If moral truths are non-causal, then no mechanism can be sensitive to the moral truths in the causal sense of sensitive. An alternative is that sensitivity requires merely that the beliefs line up well with the moral truths. More precisely, the beliefs counterfactually track the relevant truths. The classic formulation of sensitivity originates in Nozick's (1981, 172) theory of knowledge. As Nozick already noticed, sensitivity must be relative to a belief forming method to escape some known counterexamples. Therefore, we get the following:

Counterfactual tracking: Your belief that p is insensitive to the relevant truth iff: if p were false and you were to form a belief about p using the same method, then you would have still believed that p.

According to Nozick, counterfactual tracking is a necessary condition for knowledge. Here, I am not concerned with knowledge, but rather with justified belief. While I do not believe that counterfactual tracking is a necessary condition for epistemic justification (neither did Nozick), I am attracted to the view that if you learn that a belief of yours is insensitive to the relevant truth, that counts against your epistemic justification for that belief. Notice that counterfactual tracking is a weaker condition than appropriate causal connection. A belief may counterfactually track the truth because of an appropriate causal relationship, but that need not be the case. It is possible for counterfactual tracking to exist for other reasons. This feature is important to the debunking debates, as we will soon see.

How can we tell whether a belief of ours counterfactually tracks the relevant truths? The first thing we need to ask ourselves is whether in this context we are allowed to assume in the outset that our beliefs are true. If we cannot rely on our judgments, there is nothing else to rely on. We should therefore believe that most of our (positive) moral beliefs are false. Why? Because once we set aside our actual moral beliefs, there are infinite epistemically possible (we can rule out obviously inconsistent sets) sets of normative claims that have similar claim to being the moral truths. If we have no way to attribute higher probability to one over the other, we should attribute equal probability to all such sets. It follows that the probability of our particular set of moral beliefs being the truths is exceedingly low. And if most of our moral beliefs are false, then we do not track the moral truths.

---

[25] Schafer (2010) pursues this line in defending moral realism from an insensitivity based argument.

What happens if it is reasonable to assume that our actual beliefs are true? Let us first think about the set of moral beliefs that we take to correspond to the pure moral facts. Some think of such facts as necessarily true. If they are, then for any such belief, the counterfactual "If p were false, then you would not have believed that p" has an impossible antecedent. It is common to call such counterfactuals *counterpossibles*. The semantics of counterpossibles are a controversial issue. According to David Lewis's very influential view, all counterpossibles are vacuously true. (I call this view *positive vacuism*. See Baras (2019, n. 2)). If Lewis is correct, and we assume that some belief of ours p is purely moral and true, then we should conclude that our belief that p satisfies counterfactual tracking. Science, as far as I can tell, is not needed to determine any of the premises that support this conclusion.

You may think however, as I am inclined to think, that counterpossibles are non-vacuous, that is, that they are neither all vacuously true nor all vacuously false. Rather, some are true and some are false depending on the relationship between the antecedent and the consequent. (For vacuists, the consequent has no influence on the truth-value of the counterpossible). Or you may just think that pure moral facts have a different kind of necessity, not metaphysical necessity (Fine 2002; Rosen *forthcoming*), or you only accept vacuism when the antecedent is logically impossible, as opposed to any other kind of metaphysical necessity (Kment 2014, 25; 220; Field 1996, 375). If you are prone to think in one of these ways, you should conclude that we do not counterfactually track the moral truths, at least with regard to pure moral truths. Why? Because if you accept metaphysical autonomy, then you should think that the nearest (impossible) worlds in which a given pure moral truth is false, is one which is only morally different from our world, not different in any non-normative respect. And since our beliefs are non-normative respects of our world, in such a world we would have the exact same beliefs as in the actual world. Let us demonstrate using an example. Consider your belief that it is bad to cause pain. If, per impossible, it were not the case that causing pain is bad, would you still believe that causing pain is bad? The answer is yes. Because the badness of causing pain is not among the causes of your belief that causing pain is bad. Nor would any of the causes of your belief change in worlds in which causing pain is not bad and everything else is like the actual world. Because, given metaphysical autonomy, such moral differences imply no non-normative differences. So, whatever causes you to believe that causing pain is bad, would still cause you to have this belief in an impossible world in which causing pain is not bad.

If this is true of pure moral facts, and impure moral facts are all partially grounded in fundamental moral facts, then it seems that if we are insensitive to pure moral facts we will be insensitive to the impure moral facts which they ground.

To conclude this part of the paper, in order to determine whether we are sensitive to the moral truths (premise (2)), there are a number of questions that we will have to settle. We will have to decide what exactly it takes to count as sensitive, we will have to decide whether we should assume that our moral beliefs are true in this context, and we will have to decide how to evaluate counterpossibles. It is difficult to see how empirical details will help us solve these issues. And once we have solved these issues, we have all the information we need and there is nothing left for empirical details to settle.

For the remainder of this paper I switch to a different strategy. Several theorists have put forward thought experiments that are supposed to demonstrate how future empirical discoveries may affect the plausibility of the Conditional Debunking Conclusion. I examine two such examples and argue that they fail at this task.

4 The coin flipping hypnotist

Derek Parfit has us consider the following example:

> Suppose we discovered that we have some belief because we were hypnotized to have this belief by some hypnotist who chose at random what to cause us to believe. One example might be the belief that incest between siblings is morally wrong. If the hypnotist's flipped coin had landed the other way up, he would have caused us to believe that such incest is not wrong. If we discovered that this was how our belief was caused, we could not justifiably assume that this belief was true. (Parfit 2017, 3:287).

Parfit takes this example to demonstrate that there is at the very least a theoretical possibility that empirical research will justify a debunking conclusion. What is it precisely about the hypnotist case that gives reason to doubt the belief? One possibility, endorsed by some authors,[26] is that whenever we learn that we could have easily had a very different belief than the one we actually have, that tend to undermine our beliefs. If this is the correct diagnosis, it may be that future science will help us determine the degree of contingency of our beliefs. I believe, however, that this is a mistaken analysis. Consider a different example:

> You learn that a hypnotist could have easily hypnotized you yesterday, and if he would have done so, he would have caused you to believe that incest is morally permissible. However, the hypnotist got caught in traffic and could not make it. So you remained with your belief that incest is wrong.

In this example, it doesn't seem right to think that the justification of the belief is undermined, despite the fact that you learn that you could have easily believed otherwise. Now consider a case in which we eliminate the contingency:

> We learn that our beliefs were caused by a coin-tossing hypnotist. However, the coin is robustly deterministic. For each toss, the coin could not possibly have landed on a different side. The same is true of the hypnotist's actions. The hypnotist could not have abstained from tossing this deterministic coin, nor could he have caused you to believe anything other than what the coin determines. (And this is the case in all the nearby possible worlds).

It seems that such a discovery causes the same kind of uneasiness as Parfit's original coin tossing hypnotist. And this shows that the reason for this uneasiness is not that the belief turns out to be highly contingent.

What else could be the source of unease in Parfit's case and my non-contingent variant? An attractive alternative is that the important feature of these cases is that the belief is caused by a mechanism that is in no way connected to the relevant truth. But what does that mean? If it is just another way of saying that the belief is insensitive to the truth then we already know where that leads us. The previous section argued that future science will not be able to help us determine whether our beliefs are sensitive to the relevant truths. If Parfit's example is a genuine example of debunking, then I do not see why we do not already know that the explanation of our moral beliefs will have the relevant ingredient to produce the same debunking effect. We know that moral truths are not connected to our moral beliefs in the same way that the truths and beliefs are unconnected in the hypnotist examples. Consequently,

---

[26] See for example Barkhausen (2016).

it seems that if robust realists want to resist the debunking conclusion, they will have to resist Parfit's verdict on his example. This however is not the focus of this paper. Here I am interested in whether future science can teach us anything that we do not already know and bares on whether the debunking conclusion is true.

5 Changing circumstances

Joshua Greene offers the following example of a debunking explanation:

> Ought we condemn all incestuous behavior? … Suppose we learn (as we likely have already) that the inclination to condemn incest of all kinds is based on an emotional response whose function is to avoid producing offspring with genetic diseases…Having made this assumption, and having learned something from science, we may now conclude that we ought not condemn all incestuous behavior—an interesting normative conclusion. (Greene 2014, 712).

The argument, I submit, is initially attractive. It does seem like a demonstration that scientific details matter, and can serve as premises in genuine debunking. The form of this argument can be generalized[27]:

1. (We learn that) intuition x was formed as a response to certain features of the environment of our ancestors.
2. If (we learn that) an intuition was formed as a response to certain features of the environment of our ancestors, we should not give epistemic weight to that intuition in circumstances that lack those features.
3. Therefore, we should not give epistemic weight to intuition x regarding circumstances relevantly different from the circumstances in which it was formed.[28]

If Greene's argument is sound, then it is a way in which future empirical research can support the Conditional Debunking Conclusion. That is because, potentially, it can be applied to all of our moral intuitions, and if intuitions are our only hope, then all of our moral beliefs would be undermined. Note though that Greene's own conclusion is not the Conditional Debunking Conclusion, which has as its consequent a global debunking of all of our moral beliefs. Greene thinks that our deontological intuitions are debunked by such reasoning, but that consequentialist reasoning remains intact. This further step falls outside the domain of this paper and has been discussed by others.[29]

As attractive as the argument may seem at first, I argue that it rests on a confusion. Whether or not incest is wrong in all circumstances depends on what is wrong with incest. If incest is intrinsically wrong, then there will be something wrong about it in every circumstance. If incest is only instrumentally wrong, then (unless other outweighing considerations come into play) incest will only be wrong in circumstances in which it leads to the unwanted

---

[27] Fraser's (2014, sec. 3.1) environment condition for evolutionary debunking arguments is based on this sort of reasoning. Sauer (2018, 34) calls this kind of debunking argument "obsoleteness debunking".

[28] This negative conclusion is slightly weaker than Greene's in the quoted text. Greene makes a further inferential jump to the positive conclusion that some cases of incest should not be condemned.

[29] Greene's claim that consequentialist reasoning does not fall prey to his own argument is criticized by Tersman (2008), Berker (2009) and Kahane (2011).

consequences. These are the two possible views you might have prior to learning the scientific explanation. For each possibility, let us examine how the new scientific information should or should not lead to a revision of belief. Suppose prior to receiving the scientific explanation of our intuition against incest you were convinced that incest was *only* wrong because it can result in offspring with genetic disorders. In that case, you should have reached the conclusion that in an era of reliable contraception (or after menopause etc.) incest would cease to be wrong. No sophisticated empirical science is needed to justify this judgment.

The interesting case would be when you were at least somewhat inclined to believe initially that incest is intrinsically wrong, or at least wrong across a wider set of circumstances, and then you are driven to reject this inclination on the basis of the scientific findings. In such a case, how exactly would the scientific data get you to change your mind? Here is one way it can. You might think: The scientific information shows me that the whole point of the anti-incest intuition is to prevent genetic disorders. Thus, I should conclude that the whole point of the norm against incest is to prevent genetic disorders. However, such an inference rests on a conflation between biological teleology and normative teleology.[30] When scientists discover the evolutionary function of some mechanism, all that means is that they have discovered a feature of that mechanism that made it more likely to survive natural selection. They have not discovered any fact about whether that feature is good or bad or what point it may have *morally*.

The situation is in some ways similar to our sweet tooth. Our preference for sweet foods evolved in a low caloric environment, where this preference was to our advantage. Today, in our high caloric environment, a sweet tooth tends to cause more health problems than benefits. Learning this evolutionary explanation should not in itself affect our attitudes towards sweetness. We don't need to know about the changing circumstances to know that overconsumption of sweet foods can cause health problems, and that is a reason to abstain from sweet foods. And nothing about the evolutionary explanation changes the fact that we still find sweet foods tasty. The evolutionary explanation is helpful in explaining why we have this tendency that seems, in our current environment to cause us more harm than good. But it doesn't give us reason to change our attitudes towards sweetness, for better or for worse.

Here is another way of seeing the problem. If you think that an evolutionary function of an intuition gives you the "point" of that intuition, meaning the underlying normative reason for the content of the intuition, then this logic should lead you to think that the point of all intuitions is to proliferate your genes. That is because that is the ultimate evolutionary function of almost everything about us. I do not see what could justify a more limited inference from the science. Should we therefore conclude that the point of all of normativity is to proliferate our genes? Should we make that the point of all of our actions? That would be an absurd conclusion.

6 Conclusion

If we assume that morality is autonomous, such that moral beliefs have to be epistemically grounded, one way or another, in moral intuitions, and if we assume in addition that because of the autonomy of morality, no future good scientific explanation will tie pure moral truths to our moral intuitions, we have all that we need to figure out from the armchair whether our moral

---

[30] Priest (1997) makes a similar point about the idea of unnatural sex. Thanks to Jessica Issarow for the reference.

beliefs are thereby debunked by an argument from insensitivity. No detailed empirical findings of future scientists can be expected to make a significant difference to this issue. At least as far as the arguments considered in this paper are concerned.

References

Allman J, Woodward JF (2008) What are moral intuitions and why should we care about them? A neurobiological perspective. Philosophical Issues 18(1):164–185. https://doi.org/10.1111/j.1533-6077.2008.00143.x

Baras D (2017) Our reliability is in principle explainable. Episteme 14(2):197-211

Baras D (2019) How close are impossible worlds? A critique of Brogaard and Salerno's account of Counterpossibles. Dialectica 73(3):315–329. https://doi.org/10.1111/1746-8361.12269

Barkhausen M (2016) Reductionist moral realism and the contingency of moral evolution. Ethics 126(3):662–689. https://doi.org/10.1086/684708

Bedke MS (2014) No coincidence? in *Oxford Studies in Metaethics*, 9:102–25. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198709299.003.0005, No Coincidence?*

Bergqvist A, Cowan R (2018) Evaluative perception: introduction. In *Evaluative Perception*, edited by Anna Bergqvist and Robert Cowan. Oxford University Press. https://doi.org/10.1093/oso/9780198786054.003.0001

Berker S (2009) The normative insignificance of neuroscience. Philos Public Aff 37(4):293–329. https://doi.org/10.1111/j.1088-4963.2009.01164.x

Braddock M (2017) Debunking arguments from insensitivity. International Journal for the Study of Skepticism 7(2):91–113. https://doi.org/10.1163/22105700-20171195

Chappell RY (2017) Knowing what matters. In *Does Anything Really Matter?*, 149–68. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199653836.003.0007

Clarke-Doane J (2016) Debunking and Dispensability. In *Explanation in Ethics and Mathematics*, edited by Uri D. Leibowitz and Neil Sinclair, 23–36. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198778592.003.0002

Clarke-Doane J, Baras D (forthcoming) Modal security. Philos Phenomenol Res. https://doi.org/10.1111/phpr.12643

Cuneo T (2007) Recent faces of moral nonnaturalism. Philos Compass 26(10):850–879. https://doi.org/10.1111/j.1747-9991.2007.00102.x

Enoch D (2011) Taking morality seriously. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199579969.001.0001

Faraci D (2019) Groundwork for an Explanationist account of epistemic coincidence. Philosophers ' Imprint 19(4):1–26

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

Field H (1996) The a Prioricity of logic. Proc Aristot Soc 96:359–379

Fine K (2002) The varieties of necessity. In: Hawthorne J, Szabó-Gendler T (eds) Conceivability and possibility. Clarendon Press, Oxford, pp 253–281

Finlay S (2007) Four faces of moral realism. Philos Compass 2(6):820–849. https://doi.org/10.1111/j.1747-9991.2007.00100.x

FitzPatrick WJ (2008) Robust ethical realism, non-naturalism, and normativity. In: In *Oxford Studies in Metaethics*, 3:159–205. Oxford University Press, Oxford

Fraser B (2014) Evolutionary debunking arguments and the reliability of moral cognition. Philos Stud 168(2):457–473. https://doi.org/10.1007/s11098-013-0140-8

Greene JD (2013) Moral tribes: emotion, reason and the gap between us and Them. Penguin

Greene JD (2014) Beyond point-and-shoot morality: why cognitive (Neuro)science matters for ethics. *Ethics* 124(4):695–726. https://doi.org/10.1086/675875

Hill S (2016) From isolation to skepticism. Erkenntnis 81(3):649–668. https://doi.org/10.1007/s10670-015-9760-z

Hopster J (2018) Evolutionary arguments against moral realism: why the empirical details matter (and which ones do). Biol Philos 33:1–19. https://doi.org/10.1007/s10539-018-9652-0

Hume D (1888) A treatise of human nature. Clarendon Press

Isserow J (2019) Evolutionary hypotheses and moral skepticism. Erkenntnis 84(5):1025–1045. https://doi.org/10.1007/s10670-018-9993-8

Joyce R (2001) The myth of morality. Cambridge University Press. https://doi.org/10.1017/CBO9780511487101

Joyce R (2016) Evolution, truth-tracking, and moral skepticism. In *Essays in Moral Skepticism*, 142–58. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198754879.003.0008

Kahane G (2011) Evolutionary debunking arguments. Noûs 45(1):103–125

Kim J (1984) Concepts of Supervenience. Philos Phenomenol Res 45(2):153–176

Klenk M (2017) Old wine in new bottles. Ethical Theory Moral Pract 20(4):781–795. https://doi.org/10.1007/s10677-017-9797-y

Kment B (2014) Modality and explanatory reasoning. Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780199604685.001.0001

Korman DZ (2019) Debunking Arguments. Philos Compass 14(12) https://doi.org/10.1111/phc3.12638

Korman DZ, Locke D (forthcoming) Against minimalist responses to moral debunking arguments. In Oxford Studies in Metaethics

Levy A, Levy Y (2018) Evolutionary debunking arguments meet evolutionary science. *Philosophy and Phenomenological Research*, October https://doi.org/10.1111/phpr.12554

Lutz M (forthcoming) The reliability challenge in moral epistemology. Oxford Studies in Metaethics

Mackie JL (1977) Ethics: inventing right and wrong. Pelican Books, New York

Majors B (2007) Moral explanation. Philos Compass 2(1):1–15. https://doi.org/10.1111/j.1747-9991.2006.00049.x

May J (2018) Regard for reason in the moral mind. Oxford University Press. https://doi.org/10.1093/oso/9780198811572.001.0001

McPherson T (2011) Against quietist normative realism. Philos Stud 154(2):223–240. https://doi.org/10.1007/s11098-010-9535-y

Mogensen AL (2016) Do evolutionary debunking arguments rest on a mistake about evolutionary explanations? Philos Stud 173(7):1799–1817. https://doi.org/10.1007/s11098-015-0579-x

Moore GE (1903) Principia Ethica. Cambridge University Press, London

Nozick R (1981) Philosophical explanations. Harvard University Press

Paakkunainen H (2018) The 'just too different' objection to normative naturalism. Philos Compass 13(2):e12473. https://doi.org/10.1111/phc3.12473

Parfit D (2011) On what matters. Vol. 2. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199572816.001.0001

Parfit D (2017) On what matters. Vol. 3. Oxford: Oxford University Press

Pölzler T (2018) Moral reality and the empirical sciences. Routledge, New York

Priest G (1997) Sexual perversion. Australas J Philos 75(3):360–372. https://doi.org/10.1080/00048409712347951

Railton P (2014) The affective dog and its rational tale : intuition and attunement. Ethics 124(4):813–859. https://doi.org/10.1086/675876

Rosen G (forthcoming) What is normative necessity? In Metaphysics, Meaning and Modality: Themes from Kit Fine, edited by M. Dumitru. Oxford University Press

Rosen G (2017) Metaphysical Relations in Metaethics. In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett, 151–69. Routledge

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

Sauer H (2018) Debunking arguments in ethics. Cambridge University Press https://doi.org/10.1017/9781108529181

Schafer K (2010) Evolution and normative Scepticism. Australas J Philos 88(3):471–488. https://doi.org/10.1080/00048400903114219

Shafer-Landau R (2003) Moral realism: a Defence. Oxford University Press, Oxford

Sinnott-Armstrong W (2006) Moral skepticisms. Oxford University Press https://doi.org/10.1093/0195187725.001.0001

Sterelny K, Fraser B (2016) Evolution and moral realism. *The British Journal for the Philosophy of Science*, February https://doi.org/10.1093/bjps/axv060, axv060

Street S (2006) A Darwinian dilemma for realist theories of value. Philos Stud 127(1):109–166. https://doi.org/10.1007/s11098-005-1726-6

Street S (2008) Reply to Copp: naturalism, normativity, and the varieties of realism worth worrying about. Philosophical Issues 18(1):207–228. https://doi.org/10.1111/j.1533-6077.2008.00145.x

Street S (2015) Does anything really matter or did we just evolve to think so? in *The Norton Introduction to Philosophy*, edited by Gideon Rosen, Alex Byrne, Joshua Cohen, and Seana valentine Shiffrin, 685–93. Norton

Sturgeon NL (1984) Moral explanations. In: Copp D, Zimmerman D (eds) Morality, reason and truth. Rowman & Littlefield, Totowa, pp 49–78

Sturgeon NL (2006) Moral explanations defended. In: Dreier J (ed) Contemporary debates in moral theory. Blackwell Publishers, Oxford, pp 241–262

Tersman F (2008) The reliability of moral intuitions: a challenge from neuroscience. Australas J Philos 86(3):389–405. https://doi.org/10.1080/00048400802002010

Vavova K (2015) Evolutionary debunking of moral realism. Philos Compass 10(2):104–116. https://doi.org/10.1111/phc3.12194

Werner PJ (2016) Moral perception and the contents of experience. J Moral Philos 13(November):294–317. https://doi.org/10.1163/17455243-4681063

Werner PJ (2017) A posteriori ethical intuitionism and the problem of cognitive penetrability. Eur J Philos 25(4):1791–1809. https://doi.org/10.1111/ejop.12272

White R (2010) You Just Believe That Because…. Philos Perspect 24(1):573–615. https://doi.org/10.1111/j.1520-8583.2010.00204.x

Wielenberg EJ (2016) Ethics and evolutionary theory. Analysis. 76:502–515. https://doi.org/10.1093/analys/anw061

Woodward JF, Allman J (2007) Moral intuition: its neural substrates and normative significance. Journal of Physiology Paris 101(4–6):179–202. https://doi.org/10.1016/j.jphysparis.2007.12.003