Synthese (2021) 198:1089–1108
https://doi.org/10.1007/s11229-019-02084-1

Moral disagreement and non-moral ignorance

Nicholas Smyth[1]

Received: 28 March 2018 / Accepted: 7 January 2019 / Published online: 5 February 2019
© Springer Nature B.V. 2019

Abstract
The existence of deep and persistent moral disagreement poses a problem for a defender of moral knowledge. It seems particularly clear that a philosopher who thinks that we know a great many moral truths should explain how human populations have failed to converge on those truths. In this paper, I do two things. First, I show that the problem is more difficult than it is often taken to be, and second, I criticize a popular response, which involves claiming that many false moral beliefs are the product of nonmoral ignorance.

Keywords Moral disagreement · Meta-ethics · Moral psychology

The existence of deep and persistent disagreement in a particular domain poses a problem for a realist about that domain. Roughly, a realist thinks that a domain is comprised of a single, coherent set of stance-independent truths, and they also think that such truths are basically accessible to us.[1] It seems reasonably clear that such a theorist has to explain how human populations have failed to converge on those truths. Should no such explanation be forthcoming, they will incur a certain theoretical cost.[2]

One very popular explanatory strategy centrally involves the claim that much moral disagreement can be explained by *non-moral ignorance* on the part of one or both disagreeing parties. For example, if you and I only disagree about the permissibility of some action because we have different beliefs about its likely consequences, then we may not disagree *morally*. Rather, we may share an underlying moral principle—avoid actions with certain consequences—and our disagreement may arise simply from the fact that I think that the consequences will obtain while you do not.

[1] Technically, a realist, qua metaphysician, need not take an epistemological position. However, it is probably no accident that existing moral realists are not skeptics: the view is designed to allow us to vindicate the objective truth of our central moral beliefs, or to allow us to "take morality seriously" (Enoch 2011).

[2] For discussion see Brink (1984), Shafer-Landau (1994), Loeb (1998), Audi (2008), Enoch (2009), Fitzpatrick (2014).

✉ Nicholas Smyth
nick.a.smyth@gmail.com

[1] Fordham University, B25, Collins Hall, 441 E Fordham Rd, Bronx, NY 10458, USA

Virtually every prominent moral realist who has written on this problem has suggested that many moral disagreements are the result of differing conceptions of non-moral reality. In this paper, I will refer to this as the *non-moral ignorance* (hereafter: NMI) strategy, and I will offer a series of reasons—conceptual, historical and moral-psychological—to think that it will almost always fail to explain away the sort of moral diversity that should concern the realist. To be clear, I do *not* think that this result renders the problem of moral disagreement intractable. Rather, my principal aim is to criticize the NMI-strategy. Before proceeding to this critique, however, I want to carefully lay out the problem of moral disagreement itself. This is a very important preliminary, because the evaluation of mitigating strategies should only proceed under the most challenging version of the problem.

1 The problem of moral disagreement

It is difficult to read anthropology or history without coming away with the impression that human moral belief is, to put it somewhat mildly, divergent.[3] For example, Michel Foucault's *Discipline and Punish* contains a historical account of a publicly attended torture-execution which took place in the 18th century. To read the account is to be viscerally confronted with a somewhat unsettling fact, namely, that moral cognition has often been far less sensitive to human suffering than it is today (Foucault 1977). There is simply no other way to interpret agents who stand and cheer the slow evisceration, torture and execution of their fellow citizens. Nor is this an isolated incident, since whole civilizations have been built on foundations of systematic cruelty and oppression.[4]

Now, these facts are supposed to make trouble for the moral realist in particular. This is because, as John Mackie says while giving his oft-cited *argument from relativity*:

> [A]ctual variations in 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 (Mackie 1977).

In what follows, I won't concern myself with the exegetical question of what Mackie actually had in mind, here. Rather, I'll try to carefully construct what I take to be the strongest version of his argument.

[3] Michelle Moody-Adams argues against this impression (see Moody-Adams 1997). I won't address her arguments here, except to say that many of them are irrelevant to the problem of moral disagreement that should concern the realist. Moody-Adams spends a great deal of time emphasizing that *cultures* are not unified, clearly definable units, but no skeptic who wants to deploy disagreement-based arguments should place any weight on the notion of *cultural* disagreement. Moral diversity amongst human beings, past and present, is enough to get the argument off the ground.

[4] Here, the reader could scarcely do better than to consult histories of the Mongol empire, wherein accounts of socially sanctioned mass murder, torture and ethnic cleansing often seem to be on almost every page (Ratchnevsky 1993).

1.1 Explanation

Notice that Mackie's focus is on *explanation*, on competing hypotheses concerning the causes of moral diversity. A substantive reply to the problem of moral disagreement consists in just such an explanation. This might seem like a trivial point, but it is not. After all, a realist might be tempted to say that a certain disagreement is explained by one party's being *irrational*, and when we inquire into their conception of rationality, we might discover that it mainly involves *having the right beliefs*, in particular, having one of the very beliefs that creates the disagreement in question. Since it is very plausible that a thing does not explain itself, this is no explanation of a moral disagreement, rather, it is merely the re-description of one. Moreover, some philosophers seem to have thought that the provision of a moral theory constitutes a reply to skeptical worries arising from disagreement.[5] However, a moral theory is a theory of what makes beliefs true and false, and not an explanation for why they vary so widely.

Why must the realist provide such an explanation? I'll shortly provide a precise outline of the argument from disagreement in Sect. 1.6, but for now, I want to provide the argument with some direction. What, exactly, is the troubling conclusion that the realist should worry about?

1.2 Disagreement as epistemology, not metaphysics

Consider two interpretations of Mackie's argument. We might call these the *metaphysical* and *epistemological* interpretations:

Disagreement as metaphysics We have good reason to think that moral disagreement is deep and intractable, and the best explanation for this phenomenon is that realism is false, or that there are no mind-independent moral truths.

Disagreement as epistemology We have good reason to think that moral disagreement is deep and intractable, and the best explanation for this phenomenon is that our moral beliefs are not formed in a reliable or trustworthy fashion.

As Sarah McGrath has convincingly argued, the metaphysical version of the argument from disagreement is very weak (McGrath 2007). We might admit that there is something mysterious about persistent disagreement in a mind-independent domain, or that we should expect convergence on such truths. However, this is only a *very* weak expectation, one which can easily be overridden. After all, the realist can simply note that moral thinking in general is subject to a series of well-known distorting factors. For example, wishful thinking, ideological influence and self-interest have all played a significant determining role in moral belief throughout human history (Boyd 1988). The realist can easily conclude that it would be surprising if we did *not* observe a great deal of moral diversity, given these widely acknowledged social-psychological facts. The anti-realist's attempt to derive metaphysical conclusions from facts about human cognition is thereby shown to be very weak indeed. In fact, David Enoch seems actually surprised by how easy this strategy is:

[5] See, for example, Alan Gewirth's defense of a broadly Kantian theory in Gewirth (1994).

> The striking fact about… the argument from disagreement is that such alternative explanations are so easy to come by: Many moral matters are complex and not at all straightforward; people are the victims of any number of cognitive shortcomings… we let our interests influence our beliefs, and given that our interests differ this accounts for differences in our beliefs; we are subject to the manipulation of others, and so to the distorting effects also of their self-interests. (2009, p. 25)

Enoch thinks that this collection of hypotheses can explain away most moral disagreement, and the fact that it is so easy for him to accomplish this feat does not seem to trouble him. But we should ask: can it really be *this* easy to wave away the threat posed by moral disagreement?

Unfortunately, it is not, because the far more troubling thing about persistent disagreement is that it suggests that our beliefs are not reliably formed. In fact, one puzzling thing about Enoch's strategy is that it involves *affirming* something very close to this skeptical conclusion. I will eventually elaborate on this, but for now it is worth noting a curious fact: when realists interpret this argument along metaphysical lines, they often end up "responding" to it by offering an epistemological conclusion which is deeply skeptical all on its own. As McGrath suggests, this is because the epistemological skepticism about morality is the real problem, here (see also Doris and Plakias 2008). Disagreement worries us because it makes us wonder why we have the right to retain justified confidence in our moral beliefs, under the assumption that such beliefs are made true by mind-independent, objective facts. In order to zero in on the precise nature of the skeptical challenge, here, I'll proceed to refine it with a series of qualifications.

1.3 Non-ideal disagreement

First, I want to emphasize that I will not, in this paper, refer to the phenomenon of *idealized* moral disagreement, or to disagreement amongst *ideally* or *fully rational* agents. Such phrases are common in the literature on disagreement, but they can make for some confusion (Parfit 2011, p. II:34, see also Shafer-Landau 2003: p. 217). For example, the realist David Brink writes that "[i]t could be reasonable to expect agreement on a set of facts only if all cognizers were fully informed and fully rational and had sufficient time for deliberation." (Brink 1989: p. 199)

Now, I think that such philosophers only mean to describe cognizers operating under reasonably favorable conditions, and there can be no objection to that. But phrases such as "fully rational" can suggest something else, namely, that we are speaking about what used to be called an *ideal observer*, who was factually omniscient, maximally imaginative, fully consistent, entirely dispassionate and completely impartial (Firth 1952). Such figures produce a well-known epistemological problem: we simply have no way to know what they would believe (Williams 2009). Suppose, for example, that objective consequentialism is true. It follows that an agent who knows *all* the relevant facts about rightness and wrongness of actions knows the entire set of their consequences up until the heat-death of the universe (Lenman 2000). This is just one of many ways in which the beliefs of ideally rational agents can be entirely inaccessible

to us. To put the point another way, actual diversity amongst real human agents is, I submit, no evidence whatsoever for *or against* the hypothesis that idealized diversity will obtain (Bloomfield 2008, pp. 339–340). So, the argument cannot revolve around the question of what ideal reasoners will believe, rather, it can only concern what will happen to actual human beings under epistemically favorable conditions.

Moreover, there is another reason to avoid talk of ideal reasoners. The most powerful version of the argument from disagreement, I will now argue, must not merely make reference to disagreement *in general*. Rather, it must include the indexical claim that *we* (or, perhaps *you* or *I*) are among the disagreeing parties.

1.4 Situated disagreement

Why do *some people* disagree in some domain? Well, as we have seen, their views may be distorted by a series of epistemic flaws. Such explanations, which make reference to powerful distorting factors, are very simple and effective with respect to this abstract question. If true, they are perfectly consistent with the relevant domain containing accessible, mind-independent truths, and that means that realism is not threatened.

However, here is a subtly different question: why do *we* disagree with some other group of people in that domain? Notice, here, that the simple, generalized distortion-explanation will not do. If the realist wants to claim that the beliefs of *all parties* are subject to a powerful distorting factor, she has leapt from the philosophical frying pan into the fire, because it is highly plausible that one cannot remain justified in holding any belief while at the same time acknowledging that it is the product of a powerful distorting factor.[6] Since she is one of the people featured in the question, this realist has almost certainly responded to the problem of moral disagreement by debunking her own moral beliefs. Since (as I have already argued), the problem of disagreement is fundamentally epistemological anyway, the anti-realist advocate for the problem will legitimately claim victory, here. If realism comes at the expense of our most cherished moral beliefs, it is entirely unclear why anyone should be a realist at all.

The implication is clear: the most demanding version of the problem of disagreement is essentially indexical, or as I will say, *situated*. The realist—*qua* moralizer with firmly held moral convictions—is being forced to confront the fact that many people have contrary (i.e. apparently false) convictions. This is *not* the hypothetical claim that ideal reasoners would disagree under some circumstances, it is that claim that you and some other people *do* disagree, and that you must take them to be mistaken. Notice that here, idealizing language can tempt us away from situated versions of the argument, and therefore away from its most powerful articulation.

Interestingly, this observation also implies that the problem is worse for some agents than for others, depending on their belief-set. However, the range of moral disagreement amongst human populations, contemporary and historical, is so broad that it is unlikely that any existing realist is off the hook.[7] Moreover, any actual moral

[6] This is denied by some philosophers. I'll discuss them in Sect. 1.6.

[7] Some *historical* agents may be off the hook, but I take it that this is intuitive. Suppose that the earliest human populations shared roughly the same basic moral beliefs; it is intuitive that they would not have to confront the problem of moral disagreement.

realist who is reading this paper almost certainly holds a large number of broadly egalitarian moral convictions which have been denied by most people in most cultures throughout history.

1.5 Core moral beliefs

Notice further that situated disagreement *might* not be as threatening to the realist as it first appears. This is because the relevant disagreements might be over propositions which are relatively *peripheral* within the realist's overall moral outlook. If this is so, the realist might simply accept that he lacks justification for these particular peripheral beliefs. This is not, on its own, particularly troubling. However, it is clear that no realist will be happy with an argument that cuts the epistemic legs out from under a significant number of their most fervently held moral beliefs.

While it is hard to say just what qualifies a belief for inclusion in this set, it is clear that only disagreement over our *core* moral beliefs should be a problem. In this context, I suggest that a core belief is one which strongly motivates the acceptance of objectivist realism in the first place. For most realists who might read this paper, this includes basic egalitarian propositions concerning the wrongness of certain forms of violence and the moral irrelevance of such things as gender, race and social class. Since there is significant contemporary and historical disagreement over such propositions, the problem remains as powerful as it initially seemed to be (indeed, it is unclear why it would concern us if there *were* no disagreement over such propositions). I conclude that if a realist wishes to "reply" to the problem by accepting that they lack justification for their core moral beliefs, the non-realist may once again claim victory.

1.6 The argument

Having outlined various qualifications and refinements, let me at last state the strongest version of the argument from disagreement. Begin with three hypotheses which characterize a non-skeptical moral realist position. If I am a realist, I assume that moral propositions are:

1. not made true by anyone's attitudes or stances towards them
2. logically consistent, and
3. epistemically accessible to human beings, such that sincere human inquirers operating in an unbiased manner under favorable conditions are very likely to arrive at the truth.[8]

I observe, empirically, that

4. entire societies have disagreed with my core moral beliefs, and

[8] While this might seem too strong, it's important to remember that *as* the realist makes moral truths less accessible, they threaten to debunk their own moral beliefs. That is, even if we grant the realist the right to describe *themselves* as sincere, unbiased inquirers operating under favorable conditions, they still need to believe that such inquirers are quite likely to arrive at the truth. Otherwise, it's hard to know how they can place much confidence in the proposition that *they* have arrived at the truth.

5. sincere moral inquiry is conducted in every society.[9]

Now, I believe that

6. my core moral beliefs are true

And it follows, from (2), (4) and (7), that

7. entire societies have been wrong about important moral matters.[10]

However, taken together, (2) and (5) entail that

8. individuals who operate in an unbiased manner under favorable conditions will almost certainly converge on true moral beliefs about important matters.

The challenge is to reconcile these final two propositions, (7) and (8), in a way that preserves my *own* claim to moral knowledge.

Of course, the obvious route here is to deny that many or most people who disagree with me *have* been operating in an unbiased manner under favorable conditions. That is, in the face of this problem, I can try to explain my own *comparative cognitive reliability*. There will be some explanation of my having arrived at these important moral truths, and a complementary explanation for how so many others failed to do so. Absent any such contrastive explanation, I may have to conclude that my own cognitive success is just a lucky coincidence. Intuitively, the discovery that one's own beliefs are at best *coincidentally* true undermines one's justification for holding them (Pritchard 2008). So, in taking ourselves to have mind-independent moral knowledge we are thereby committed to showing, not only that rival moral codes are largely explained by the kinds of distorting forces Enoch cites, but also that our *own* moral code is mainly explained by the operation of comparatively trustworthy or reliable forces. The murkier this explanatory contrast turns out to be, the less confidence we may rationally place in either our core moral beliefs or in moral realism.

Now, all of this comes with an important caveat. Some philosophers actually deny the intuition that coincidentally true belief is necessarily undermined in this way. They argue that even in our disagreements with *peers*—persons whose belief-formation mechanisms are equally trustworthy—we may retain justified confidence in our beliefs. Indeed, they think that we may retain this confidence even when we *lack* dispute-independent reasons to remain confident. This is usually because they reject the idea that *higher-order evidence* ought to rationally affect your confidence in a proposition (Kelly 2005; Weatherson 2010). A moral realist might try to dismiss the problem here by simply adopting a version of this view. For example, they might pursue the strong claim that all higher-order evidence can be rationally ignored, or they might simply argue for a weaker and potentially more defensible claim: that we can safely ignore this sort of evidence when we already have a great deal of first-order evidence for our beliefs.

[9] By "sincere moral inquiry", I do not wish to imply anything like rarefied philosophical inquiry, or even consciously conducted rational reflection. Rather, I mean something more like *preparedness to respond to reasons*, where the inquirer's emotional and cognitive faculties are attuned to the possibility of newer and better moral evidence.

[10] The qualifier "important moral matters" is only meant to capture the idea that we must take our *core* more beliefs to be about important issues. It seems reasonable to say that the beliefs we most care about are those which we also take to concern deeply important matters.

While several moral realists have pursued this kind of strategy, unfortunately I cannot discuss its merits here.[11] I only want to note that given the wide array of more positive or substantive responses to the problem that have appeared over the years, it's clear that realists as a group have often preferred—perhaps only for dialectical reasons—to block the argument at an earlier stage. That is, if the realist could successfully motivate the idea that most people who deny the truth of our core moral beliefs are somehow irrational, unreliable or untrustworthy, they could avoid having to deal with several thorny epistemological questions. Thus, for my purposes, I need only note that moral realists *have* been quite interested in providing explanations for their own comparative cognitive reliability. Indeed, this is because my central topic in this paper is precisely just such an explanation—the explanation which says that one disagreeing party is under the influence of non-moral ignorance.

I conclude, then, that moral disagreement is primarily a problem for realism because it highlights a need to draw an epistemically relevant distinction between our *own* views and the large number of contrary views that populate human history. The problem of disagreement is therefore a kind of *reliability challenge*: it presses us to explain our own cognitive success as well as the cognitive failure of those who disagree with us. Now, having outlined what I take to be the strongest version of the argument, I want to proceed to my main discussion. It should now be clear that the problem is basically a posteriori in nature, since it requires us to ask difficult empirical questions about the explanation for various kinds of moral beliefs, and to determine whether those explanations favor our own beliefs over their contemporary and historical rivals. And as we have already seen, some philosophers have taken up this challenge, offering empirical explanations for moral diversity which, they think, can defuse the problem posed by moral disagreement.

Since I have criticized many of those strategies elsewhere, I take no position on them here (Smyth 2017). I do not claim that *no* such strategy can work; all we can do is examine the proposals as they come in. However, one very popular response to the problem deserves far more scrutiny than it gets. It goes like this: many human groups have indeed been mistaken in moral matters, but this is very often due to *non-moral ignorance*. This seductive thought nicely blunts the skeptical force of the argument, implying as it does that a great deal of moral diversity would evaporate amongst groups of believers who were reasonably well-informed on non-moral matters. This is the NMI (non-moral ignorance) explanation. What, exactly, does it look like?

2 Non-moral ignorance

It should be uncontroversial that moral beliefs *can* be explained, at least partly, by non-moral beliefs. When we learn that the relevant non-moral beliefs are false, this seems to cast doubt on the epistemological status of the moral beliefs in question. Moreover, it is often thought that many of the most virulent historical beliefs—for example, racist and sexist beliefs—can be explained in just this way. If he can substantiate this claim, the realist has arguably provided a response to the problem of moral disagreement, an

[11] See Enoch (2010), Wedgwood (2010), Setiya (2012).

explanation which entails that *his* beliefs are epistemologically superior to those of his historical ancestors.

Admittedly, the explanation enjoys a certain superficial plausibility. After all, to take one example, many sexists have openly denied that women and men are intellectual equals. Similarly, it is not hard to go into histories of racism—anti-African and anti-Semitic racism in particular—and find various bigots insisting that some the members of some hated group are somehow less than human. Since these beliefs are manifestly false, it seems clear that the explanatory burden has been discharged, and that a realist can safely dismiss the skeptical threat that such moral views might seem to pose. In particular, this is because the strategy often seems to reveal underlying moral *agreement*: the thought is that most people have shared the fundamental belief "all humans are morally equal", and that their *moral* beliefs about certain types of people have been unfortunately distorted by false empirical beliefs concerning their sub-humanity.

It is hard to exaggerate the popularity of the NMI-hypothesis amongst those who want to explain away moral disagreement; the literature on this topic is chock full of confident support for it.[12] Consider this passage from Ralph Wedgewood:

> Aristotle argues for the view that certain forms of slavery are perfectly just. On this point, Aristotle disagrees with most contemporary Western thinkers. But it is clear that Aristotle's view on this point rests on certain non-moral beliefs. Specifically, his view rests on his non-moral belief that non-Greeks are mentally inferior to Greeks to such a degree that it is actually in the interests of non-Greeks to live as the slaves (that is, the 'human tools) of Greeks. As it happens, these non-moral beliefs are false, and have been decisively refuted by empirical investigations that are accessible to all educated people today (Wedgwood 2014, p. 25).

Here, we can see all of the elements of a typical NMI-explanation. Wedgewood thinks that Aristotle's pro-slavery beliefs are *obviously* explained by his non-moral beliefs, and that since such beliefs are decisively refuted, our disagreement with Aristotle on this point has no skeptical implications whatsoever. This is, again, because the explanation reveals an underlying moral *agreement*: both we and Aristotle are (according to this line of thought) committed to the view that more intelligent beings should control the lives of creatures who are below a certain intelligence-threshold, if only because the latter kind of being is incapable of securing its own well-being. Importantly, like many philosophers enamored with this mitigating strategy, Wedgewood offers no supporting evidence for this explanation whatsoever, presumably because he takes it to be so obviously correct in this case.

[12] For example:

Many genuine moral disagreements depend on disagreements over the nonmoral facts (Brink 1989)

Careful philosophical examination will reveal, I believe, that agreement on nonmoral issues would eliminate almost all disagreement about the sorts of moral issues which arise in ordinary moral practice (Boyd 1988).

Much moral disagreement stems either from disagreement about what the relevant nonmoral facts are, or is due to some error of instrumental reasoning. Clearing up these errors and getting consensus on non-moral facts would remove a great deal of moral disagreement (Shafer-Landau 2003).

I now want to argue against this impression. I think that this is almost always a bad explanation of moral disagreement, and that the NMI-hypothesis will decisively fail with respect to virtually all of the moral diversity that should concern us. To be clear, I do not wish to prove that *all* NMI-explanations will fail. However, I think that there are several reasons to think that any given NMI-explanation will almost certainly prove unsatisfactory.

2.1 Covertly moral concepts

The first problem can be motivated by reviewing a well-known difficulty which affects discussions of the permissibility of abortion. Since all sides to this moral dispute seem to agree that it is seriously wrong to kill a person, some have been tempted to think that this debate might be resolved if the warring parties could simply agree on a definition of the non-moral concept *personhood*. However, as is often pointed out, the concept *person* is not morally neutral. Rather, its application is usually sensitive to the speaker's antecedent moral views, and it is a mistake to think that the content of the judgment "it's as person" can be represented in some morally neutral way. This means that disagreement over abortion is *not* traceable to a "non-moral" one over the nature of personhood, because *person* is a covertly moral concept. Such concepts play an important role in moral discourse, but not by having a fixed, universally shared definition. Rather, they normally function as markers of moral significance, and a great deal of moral discussion involves what David Plunkett calls *meta-linguistic negotiation* over how such terms should be used (Plunkett and Sundell 2013).

When a philosopher wants to deploy the NMI-explanation, they must take care to ensure that the allegedly 'non-moral' judgment in question does not deploy one of these concepts. This is because such judgments are in fact *moral* ones, and the NMI explanation will collapse into an explanation which explains moral diversity by citing moral diversity. This, in turn, will lead us straight back into the clutches of the skeptical argument from disagreement.

In fact, this is precisely what happens with many NMI-explanations of racist belief. Take, for example, the idea that such beliefs can be explained by citing false beliefs about the *humanity* of their targets. It is true that the phrase "sub-human" and its cognates is easy to find in the annals of history. But if the concept *human* is like the concept *person*, then the NMI strategy seriously mischaracterizes the nature of such claims as "X is sub-human". And, as you may have guessed, I think that the concept *human* is a covertly moral concept. Doubtless it can be defined in purely non-moral terms, but that is not relevant, here. What is relevant is that when various moralizers, past and present, deny humanity to various groups, *they* have not been operating under any such definition. Rather, I think that "human" and its cognates have served as markers of moral significance, standing for highly valued traits of character.

We can see this when we refer to actual historical texts, which, while often disturbing, offer real insight into the nature of such concepts. Here, for example, is the French abbot Peter the Venerable, writing in the Twelfth century:

> Yes, you Jews. I say, do I address you; you, who till this very day, deny the Son of God. How long, poor wretches, will ye not believe the truth? Truly I doubt

> whether a Jew can be really human (*homo*)… I lead out from its den a monstrous animal, and show it as a laughing stock in the amphitheater of the world, in the sight of all the people. I bring thee forward, thou Jew, thou brute beast, in the sight of all men.[13]

It is not hard to see that in Peter's mouth, the Latin word *homo* does not call forth a concept with purely non-moral application-conditions. Jewish people are called inhuman because they commit what was thought to be a fundamental sin when they denied the son of God, and not because they lack any particular physical or mental feature. Indeed, since Peter explicitly means to *address* Jewish people, he cannot really think that they lack the basic intellectual capacities which distinguish us from animals, and that is part of what makes the passage (and many others like it) so puzzling. Notice that the only genuinely "non-moral" belief that is operative here is straightforwardly *true*, since a Jewish person, almost by definition, denies that Jesus was divine. Second, our disagreement with Peter over the moral status of Jewish people bottoms out, not in some disagreement over non-moral facts, but rather in conflicting views about the moral status of those who make a certain theological claim. In other words, the NMI strategy collapses, revealing a good old-fashioned moral disagreement.

Nor is Peter an isolated case. As David Livingstone-Smith has argued, denials of *humanity* to various beings almost always presuppose substantive and contested moral views.

> The notion of subhumanity—the idea that other organisms are less or lower than human beings in a specifically moral sense—presupposes the idea of a moral hierarchy. This idea was traditionally represented by the Great Chain of Being… [w]e human beings placed ourselves just "a little lower than the angels"… and assigned every other organism to one or another lower rank. (Smith 2014).[14]

In sum, disagreements about who is and is not human are ordinarily moral disagreements disguised as non-moral ones; *human* is, for most people, a thick ethical concept.[15] So the advocate for the NMI-strategy has to ensure that a given "non-moral" disagreement really *is* a non-moral disagreement, or that it is not a moral disagreement in disguise. Moreover, even if this is established, notice that the primary task still remains before them. They must give us good reasons to think that the relevant non-moral disagreement *actually explains* the moral disagreement. This, as it turns out, is going to be a lot harder than it looks. The main issue becomes clear when we look at some contemporary moral psychology.

[13] As quoted in Brown (1992)

[14] Importantly, Livingstone-Smith anticipates and rejects the notion that contemporary people have abandoned this conceptual presupposition, noting that "the idea of a normative hierarchy is still very much alive in our moral psychology… [w]e regard our own kind as having the greatest value, and think of animals as having greater value than plants. We esteem "higher" animals like primates more than "lower" animals like invertebrates… terms like "higher" and "lower," which roll off the tongue so easily, are hierarchical and ultimately normative notions that are inconsistent with a scientific conception of the biosphere.".

[15] Moreover, as psychologists continue to discover that more and more of our apparently "factual" concepts are heavily moralized—that we are, in Joshua Knobe's words, "moralizing creatures through and through"—the range of disagreements that an NMI-strategist can even *in principle* explain might shrink even further (Knobe 2010).

2.2 Post-hoc rationalization

One of the most important general lessons of modern psychology is this: human agents very often lack direct insight into the real explanation for their own mental states. In particular, when they are asked why they believe something, they are inclined to cite rationalizing lines of thought which actually play little to no role in the germination of their belief. While this phenomenon is by no means a *discovery* of modern thought, it received heavy emphasis in the work of modern philosophers like Schopenhauer, and, perhaps most notably, Nietzsche.

Here is Nietzsche, writing in his favored aphoristic style:

> "I did that" says my memory. "I couldn't have done that" – says my pride, and stands its ground. Finally, memory yields. (Nietzsche 1886/1990, p. 68)

The aphorism highlights a phenomenon that is central to Nietzschean moral psychology, that of motivated cognition (Kunda 1990; Kruglanski 1996; Kundra and Sinclair 1999; Murray 1999). Nietzsche believed that each of us is a bubbling cauldron of drives and desires, and that each drive seeks to express itself by influencing both action and belief. He delighted in revealing that some putatively 'cognitive' process was at bottom explained by the operation of such drives. For example, he claimed that moral theorists

> take a conjecture, a whim, an "inspiration" or, more typically, some fervent wish that they have sifted through and made properly abstract – and they defend it with rationalizations after the fact. They are all advocates who do not want to be seen as such. (Nietzsche 1886/1990, p. 5)

While these theorists may cite a complex chain of reasoning in support of their beliefs, Nietzsche suggests that the "fervent wish" plays the leading explanatory role, here. He is thus a historical progenitor of a contemporary school of moral psychology which suggests that in effect we are *all* like this with respect to most of our moral beliefs. Most prominent in this school is Jonathan Haidt, whose empirical work strongly suggests that the reasons we tend to cite in support of our moral beliefs very often do not actually explain those beliefs. Rather, the explanatory arrow runs in the other direction, since people will very often cling to their moral beliefs even if they are convinced that their factual rationalizations are false (Haidt 2001; Kelly et al. 2010; Kahneman 2011).

Two things must be said about this psychological theory. First, if it is true, nothing at all follows about the epistemological status of our moral beliefs. While many are inclined to suspect that Haidt's work debunks a great deal of moral belief all on its own, this is simply not so. After all, our deep desires, intuitions or wishes might be tracking the moral truth, or properly sensitive to moral facts (Sauer 2012). It is only if we accept a substantive (and highly controversial) *normative* claim—that moral beliefs are only justified if they are principally explained by conscious cognition—that any debunking conclusion follows. Not only is this view independently dubious, it is also completely inessential to the moral-realist package.

Second, I do not wish to claim that dumbfounding is universal or that every moral belief is best explained in this way. It is enough for my purposes that many or most

moral beliefs can be explained by this sort of emotion or intuition (I will say why this is so once the problem has been made clear).

Even in advance of detailed empirical investigation, this weaker hypothesis has enormous plausibility: any keen observer of human social interactions cannot fail to notice that moral beliefs are intricately bound up with deep emotional responses, and that moral discussions very often consist in a series of frantic attempts to rationalize these responses in just the way that Haidt and Nietzsche describe. This, in turn, reveals something very important about the relation between moral beliefs and commonly articulated factual justifications for those beliefs.

The point can be easily made with an example. Suppose I say to you:

(N) "Cockroaches are awful creatures, and we ought to exterminate them."

You disagree, and ask me to defend these normative beliefs. I respond:

(F) "Well, they spread disease."

Here is what *is* true: as a purely logical matter, (F) can help to rationalize (N). In other words, there exists a valid argument which uses (F) as a premise and which terminates in (N). That argument looks like this:

(N2) Anything which spreads disease is awful and ought to be exterminated.
(F) Cockroaches spread disease.

Therefore (N) cockroaches are awful and ought to be exterminated.[16]

Indeed, in uttering (F), I strongly implicate that I have just such an argument in mind. But the fact that this rationalizing argument exists does not mean that my belief (F) *explains* (N).[17] In fact, in this particular case, that explanatory hypothesis is implausible: if I am at all typical, my belief (N) is probably an expression of disgust, and (F) is a post hoc rationalization of (N). If you show me good evidence that (F) is false, I will probably cling to (N) anyway.

I conclude that even when historical agents have articulated false non-moral justifications for their moral beliefs, this is only very weak evidence that those factual beliefs *explain* the moral ones. But this, in turn, suggests that the NMI-strategy will fail to explain away a huge number of moral disagreements. How many, exactly?

Haidt and his colleagues claim that their model describes "the great majority of moral judgments made by the great majority of people," and that judgments actually produced by factual reasoning are "rare, occurring primarily in cases in which the initial intuition is weak" (Haidt and Bjorklund 2008, p. 193). For illustrative purposes, we might adopt a somewhat simplistic statistical approach and ask what happens when we conservatively assume that just 75% of human moral belief is explained in the way that Nietzsche and Haidt suggest. For any given pair of beliefs, we may infer that there is a $(.75 \times .75) = 56\%$ chance that they are *both* explained by gut intuition, and that an

[16] Of course, an argument which reflected the actual thinking of cockroach-haters would include a number of further background assumptions about cockroaches: that they serve no useful purpose, that they feel no pain, etc. I omit these details only for the sake of clarity. I thank an anonymous reviewer for encouraging me to be clearer about this.

[17] Nor does it even entail that I believe (N2), after all, I might just be saying this because I think that *you* believe (N2).

NMI-explanation will fail to explain away a disagreement involving those two beliefs. It also implies that only $(.25 \times .25) = 6\%$ chance that any pair of moral beliefs is primarily explained by non-moral beliefs, and thus that only 6% of moral disagreements can be *definitively* explained away via the invocation of an NMI-explanation.[18] But this merely statistical analysis is doubly conservative: after all, as Haidt implies, the model predicts that our strongest, most firmly held moral intuitions (concerning basic equality or human dignity, for example) are much more likely to influence our moral judgments, and it is plainly *these* which normally explain our core moral beliefs.[19] Since disagreement over core belief is what should concern the realist, it is very likely that Haidt's model entails that NMI-explanations will fail much more often for the cases that matter.

Merely statistical analyses aside, this problem can be made vivid when we consider key moral disagreements in a contemporary context. While she does not explicitly endorse this explanation, Katia Vavova has argued that disagreements over the morality of capital punishment could be explained by divergent views on its effectiveness as a deterrent:

> [M]any apparent moral disagreements aren't really disagreements about moral matters. For example, we might disagree about whether the death penalty is justified because we disagree about whether it has deterrent effects. But whether the death penalty has deterrent effects is an empirical matter. Our disagreement, then, isn't a moral disagreement (Vavova 2014).

Again, Vavova isn't explicitly defending the explanation itself in this passage. But the ease with which this particular NMI-hypothesis is offered is, I think, revelatory of a general tendency to forget about moral psychology. No matter what your views on capital punishment, I'm willing to bet that a little honest reflection will reveal that facts about deterrence are not particularly explanatory. An easy way to see this is to ask whether you would change your views if you learned that the other side was right about on-balance deterrence.[20] Each of us, I think, tacitly understands that retributive emotions play a key role in generating support for the practice, and that emotions such as compassion probably undergird much opposition to it. Indeed, This is exactly what most available research on this question reveals (Ellsworth and Gross 1994; Lambert et al. 2004). Alleged facts about deterrence are often invoked in order to rationalize support or opposition, but this is very often an exercise in post hoc rationalization.

[18] I use the term "definitively" because the middle cases—where only one disagreeing party has the required sort of belief—introduce complexities that I cannot fully address here. Roughly: NMI hypotheses *can* explain away those disagreements, but they will not always do so.

[19] Indeed, this may be something very near to a definitional truth, since it is hard to know how something could count as a "core" moral belief unless it were deeply entwined with emotional commitment. I believe that there's nothing morally wrong with spitting on the street, and someone might disagree. But I would not leap to a defense of my view or be particularly troubled by the prospect of dropping it. Plausibly, the reason I would respond differently to a disagreement over the moral status of *slavery* is that my moral belief is deeply rooted in my emotional and intuitive sensibility.

[20] Of course, there are extremely silly views about the deterrence effect of capital punishment which might move us if we held them, for example, the view that a single execution will prevent all future violent crime. But this is not relevant to the present debate, since both pro- and anti-capital punishment advocates reject this empirical claim.

So, while it is true that the following argument exists:

(N2) Practices which significantly deter violent crime ought to be adopted.
(F) The practice of capital punishment significantly deters violent crime.

Therefore, (N) we ought to adopt the practice of capital punishment.

This does not make it at all likely that agreement on its effectiveness as a deterrent would significantly reduce disagreement over the moral status of capital punishment. The alternative is far more likely: just as our cockroach-hater would continue to advocate for the extermination of cockroaches, most supporters of the death penalty would continue to insist that those who commit the most heinous crimes deserve to die, even if convinced that their deaths had no deterring effects.

In sum, if there is a genuinely non-moral disagreement that could explain a moral disagreement, and even if the disagreeing agents *explicitly avow* the non-moral beliefs in question when asked to justify their moral beliefs, this is at best very weak evidence that the non-moral beliefs actually do the explanatory work that the NMI-advocate needs them to do. In my view, all of this means that moral philosophers should be far less confident in the explanatory power of non-moral ignorance. Having outlined what I take to be three very serious problems for the NMI-strategy, I want to close by briefly discussing an objection which my analysis might seem to invite.

3 Disagreement and reflective equilibrium

At this point, a philosopher sympathetic to realism might be wondering why I have spent so much time defending a psychological model which *itself* forms the basis of a powerful reply to the problem of moral disagreement. That problem, after all, requires us to explain why human populations have diverged so widely on moral matters, in particular by portraying a great deal of moral belief as the product of distorting or epistemically untrustworthy forces. Now, recall that quite a few moral realists accept the Rawlsian epistemological model of *wide reflective equilibrium*, which portrays justified moral belief to be the product of a complex process of reflective adjustment (Daniels 1979; Brink 1989; Sayre-McCord 1996; Scanlon 2014). If Haidt and Nietzsche are right, then most moral beliefs simply do not have this explanatory history. This, in combination with Rawlsian moral epistemology, implies that most moral beliefs are epistemically unjustified, and helps to explain wide moral diversity. So why does the failure of the NMI-strategy matter? Can't the moral realist, at this point, simply declare that the problem of moral disagreement has been solved? As Simon Fitzpatrick suggests:

> [moral disagreements] are culturally entrenched as a result of the fact that we, as humans, automatically and unconsciously internalize the prevailing norms of our social group, rarely, if ever, reflect on the judgments these internalized norms give rise to, and thus rarely bring them into contact with the rest of our beliefs. This leaves open the possibility that such disagreements may be resolved were the parties to the disagreement to consciously reflect on their brute moral

> intuitions, bring them into contact with their non-moral beliefs, and engage in the pursuit of wide reflective equilibrium (Fitzpatrick 2014, p. 187).

First, a caveat: I want to stress that I have taken no position on whether the realist *can* successfully reply to the problem of moral disagreement. In fact, my suspicion is that it can be done, and that it will require a wholly new kind of moral epistemology.[21] However, I do want to argue that Fitzpatrick's strategy is not going to work.

First, recall that the most powerful version of the problem of moral disagreement is *indexical*, in the sense that it asks the realist to defend *their* core moral beliefs from the skeptical threat posed by people who reject them. This is why I took care to emphasize that a story which portrays human moral belief *in general* as being subject to distorting forces is no answer to the problem. Rather, what we need is a story which portrays the realist's core beliefs in particular as being the product of their favored, epistemologically superior process in this case, something resembling the method of reflective equilibrium. Here, there are very serious grounds for suspicion: it is not at all clear why we should think that (for example) anyone's opposition to slavery has the explanatory history outlined by Rawls. Not only is the method of reflective equilibrium a fantastically rarefied, remote procedure, the Haidt–Nietzsche model—which we are accepting here for the sake of argument—entails that almost every such self-description will be an exercise in self-deception. This strategy, like Enoch's, runs the risk of debunking the realist's core moral beliefs, by setting the epistemological bar too high.

Moreover, there is an equally serious reason to worry, here, a reason which derives from the many ways in which human belief-formation and transmission is profoundly *social* (Sripada and Stich 2006; Doris 2015, pp. 103–110).[22] As Fitzpatrick himself admits, we begin our mature adult lives with a set of intuitions and emotions bequeathed to us by our cultural heritage. This occurs via a complex process involving parents, peer groups, educators and symbolic forms created by cultural innovators, and these influences continue to work on us for the entirety of our natural lives. According to the realist who wishes to deploy Fitzpatrick's strategy, virtually all of these people have been forming their moral beliefs in an untrustworthy fashion. If the Rawlsian model of reflective equilibrium is a necessary condition on epistemic justification in the moral domain, then almost all moral beliefs in any culture will be unjustified. This is because they will in general be the product of unreflective intuition which is merely rationalized ad hoc—precisely the sort of belief-formation that the method of reflective equilibrium is supposed to prohibit, according to Fitzpatrick.

The method of reflective equilibrium asks us to begin with (and give considerable weight to) our *considered* moral judgments, those which equilibrium-theorist David Brink describes as being formed under conditions of "general cognitive reliability".[23]

[21] Here, I am entirely sympathetic to Karen Jones' discussion in Jones (2005). She suggests that any coherence-method in ethics must adjust its principles in accordance with the huge amount of information about morality that is being produced by the social, historical and anthropological sciences.

[22] Indeed, Haidt himself has despaired at the fact that critics of his theory have neglected its deeply social character (Haidt and Bjorklund 2008).

[23] For Brink, such a belief is "well informed… results from good inference patterns… is not distorted by obvious forms of prejudice or self-interest… held with some confidence, and is relatively stable over time." (Brink 1989, p. 132).

Yet, it should be uncontroversial that an initial set of firmly-held moral convictions is not formed under reliable conditions if it is the product of an epistemologically bankrupt moral culture. So far as I can tell, this strategy, in combination with facts about the manner in which moral beliefs are formed and transmitted, renders it unlikely that we will be permitted to allow *any* of our most basic moral intuitions to play a role in our moral deliberation. But this, again, is just the sort of skeptical result we were trying to avoid.

Assuming that they do not wish to revert to the strategy mentioned in Sect. 1.6—that of simply rejecting the significance of higher-order evidence arising from disagreement—this realist now has to insist that they have already suspended and escaped pervasive social influence by developing cognitively reliable moral judgments in a manner that is insulated from the negative influence of their moral culture. Of course, anyone can *say* this. But in order to justifiably assert such claims, the realist needs to tell us why we shouldn't be enormously suspicious here, *especially* given the fact that after (allegedly) suspending a myriad of social influences, they have somehow arrived at the very same set of broadly egalitarian moral convictions that one would *expect* them to have arrived at, given their social and historical context.[24] I don't think that any such claim must be illegitimate, but the evidentiary ball is firmly in the realist's court, here.

Of course, we should recall that even Haidt allows conscious reasoning to play a role in belief-formation, particularly when the agent is made explicitly aware of non-conscious or social influences on their beliefs (Kennett and Fine 2009). Here, the realist might simply accept that their beliefs aren't in particularly good shape at present, but rest easy with a conjecture: anyone who sincerely and effectively engaged in wide reflective equilibrium *would* converge on the view that slavery is wrong, that people are intrinsically equal, and so on. But this, I think, is to admit defeat. For no skeptic should ever accept a pure hypothetical in response to an epistemological challenge which provides positive evidence of unreliability.[25] If we show someone that a red light is shining on an object, they do not retain any justification for their belief that the object is red by citing the fact that people with good vision *might* see it as red in normal viewing conditions. What we need is positive evidence that this is so. This sort of evidence might be forthcoming, but at present I don't think we have any.

4 Conclusion

In this paper, I have tried to do two things. First, I have tried to give the argument from moral disagreement in its strongest form. I have claimed that it is best interpreted as an *ad hominem* epistemological challenge launched from within a first-order moral perspective. Second, I have tried to show that one popular strategy for dealing with

[24] I should stress that this is consistent with their rejecting a great deal of contemporary moral opinion. All I want to say is that virtually every existing realist moral philosophy is committed to a set of very basic propositions concerning the fundamental equality of persons. This is our contemporary moral legacy, and it is not shared by a huge number of historical cultures.

[25] Skeptical arguments which merely cite the *possibility* of our being massively mistaken in some domain may be more vulnerable to this reply.

that problem is far less promising than others have taken it to be. This is for at least two reasons:

1. There is good reason to think that many apparently non-moral disagreements are in fact moral disagreements in disguise.
2. There is good reason to think that even genuinely non-moral disagreements will very often fail to explain moral disagreements, since the relevant non-moral beliefs are very often post hoc rationalizations for deeper, more explanatorily basic moral attitudes.

This pair of possibilities dramatically reduces the potency of the NMI-strategy. Moreover, I should stress that these are not *mere* possibilities. Unlike many NMI-strategists, I have not satisfied myself with just-so stories.[26] Rather, I have tried to cite concrete historical and psychological evidence in favor of my two main claims. Of course, I cannot provide decisive evidence for these hypotheses, but some evidence is better than virtually no evidence at all.

Finally, I'll end with an unexpected twist: the arguments in this paper do not just cut against the moral realist. Importantly, several anti-realist moral relativists—who index moral truth to either individual or social perspectives—have *also* tried to explain away a great deal of moral disagreement by deploying the NMI-strategy. It is perhaps surprising that they share a strategy with the realist, but it should not be. After all, if subjectivist or cultural relativist theories are correct, the existence of a wide variety of moral perspectives might seem to entail that a great many abhorrent, evil or callous moral beliefs are in fact *true*. Support for the institution of slavery, for example, is virtually a cultural universal before 1700, but most contemporary relativists do not want to admit that huge numbers of agents have *truly* believed in their right to own other human beings.

In order to avoid this result, the relativist opens up a gap between the actual moral beliefs of individuals or cultures and idealized truth-grounding beliefs, the ones that agents or groups *would* have if their belief-formation processes were cleaned up in various ways. And a very popular way of doing this is to claim that an agent's beliefs or attitudes can only count as part of a truth-grounding moral perspective if those beliefs or attitudes are not essentially explained by non-moral ignorance (Harman 1975; Street 2009, p. 281). This, of course, is just the NMI-strategy. If my argument in this paper is right, the strategy is equally unavailable to the relativist, who must find some other defusing explanation of her own.

In other words, moral disagreement is a problem for *anyone* who thinks that they possess moral knowledge which large numbers of other human agents have lacked. It is a problem for the realist because it threatens to debunk their actual moral beliefs, and it is a problem for the anti-realist because it might force them to say that many morally abhorrent beliefs have been (and are) true. Such theorists may continue to try to explain away a great deal of moral disagreement, but in my view, they cannot reasonably do so by citing the influence of non-moral ignorance.

[26] For example, Brink's long discussion (1989) of the "distorting" factors which might explain false moral belief contains no historical or psychological detail.

Acknowledgements The author wishes to thank Robert Joynt, Miquel Miralbes del Pino, and Iain Laidley for a great deal of very valuable discussion.

References

Audi, R. (2008). Rational disagreement as a challenge to practical ethics and moral theory: An essay in moral epistemology. In Q. Smith (Ed.), *Epistemology: New essays*. Oxford: Oxford University Press.

Boyd, R. (1988). How to be a moral realist. In G. Sayre-McCord (Ed.), *Essays on moral realism* (pp. 181–228). Ithaca: Cornell University Press.

Bloomfield, P. (2008). Disagreement about disagreement. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol. 2. The cognitive science of morality: Intuition and diversity* (pp. 303–331). Cambridge, MA: MIT Press.

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

Brink, D. (1989). *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.

Brown, M. L. (1992). *Our hands are stained with blood*. Shippensburg: Destiny Image Publishers.

Daniels, N. (1979). Wide reflective equilibrium and theory acceptance in ethics. *Journal of Philosophy, 76*(5), 256–282.

Doris, J. M. (2015). *Talking to our selves: Reflection, ignorance, and agency*. Oxford: Oxford University Press.

Doris, J. M., & Plakias, A. (2008). How to argue about disagreement: Evaluative diversity and moral realism. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol. 2. The cognitive science of morality: Intuition and diversity* (pp. 303–331). Cambridge, MA: MIT Press.

Ellsworth, P. C., & Gross, S. R. (1994). Hardening of the attitudes: Americans' views on the death penalty. *Journal of Social Issues, 50*(2), 19–52.

Enoch, D. (2009). How is moral disagreement a problem for realism? *Journal of Ethics, 13*(1), 15–50.

Enoch, D. (2010). Not just a truthometer: Taking oneself seriously (but not too seriously) in cases of peer disagreement. *Mind, 119*(476), 953–997.

Enoch, D. (2011). *Taking morality seriously: A defense of robust realism*. Oxford University Press.

Firth, R. (1952). Ethical absolutism and the ideal observer. *Philosophy and Phenomenological Research*, *12*(3), 317–345.

Fitzpatrick, S. (2014). Moral realism, moral disagreement, and moral psychology. *Philosophical Papers, 43*(2), 161–190.

Foucault, M. (1977). *Discipline and punish: The birth of the prison*. New York: Vintage.

Gewirth, A. (1994). Is cultural pluralism relevant to moral knowledge? *Social Philosophy and Policy, 11*(1), 22–43.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814.

Haidt, J., & Bjorklund, F. (2008). Social intuitionists answer six questions about morality. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (Vol. 2). Cambridge: MIT Press.

Harman, G. (1975). Moral relativism defended. *Philosophical Review, 84*(1), 3–22.

Jones, K. (2005). Moral epistemology. In F. Jackson & M. Smith (Eds.), *The Oxford handbook of contemporary philosophy*. Oxford: Oxford University Press.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Macmillan.

Kelly, T. (2005). *The epistemic significance of disagreement* (Vol. 1, pp. 167–196)., Oxford studies in epistemology Oxford: Oxford University Press.

Kelly, D., Faucher, L., & Machery, E. (2010). Getting rid of racism: Assessing three proposals in light of psychological evidence. *Journal of Social Philosophy, 41*(3), 293–322.

Kennett, J., & Fine, C. (2009). Will the real moral judgment please stand up? *Ethical Theory and Moral Practice, 12*(1), 77–96.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences, 33*(4), 315.

Kruglanski, A. W. (1996). Motivated social cognition: Principles of the interface. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles*. New York: Guilford Press.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480.

Kundra, Z., & Sinclair, L. (1999). Motivated reasoning with stereotypes: Activation, application, and inhibition. *Psychological Inquiry, 10*(1), 12–22.

Lambert, E. G., Clarke, A., & Lambert, J. (2004). Reasons for supporting and opposing capital punishment in the USA, Toronto. *Internet Journal of Criminology, 1,* 1–34.
Lenman, J. (2000). Consequentialism and cluelessness. *Philosophy and Public Affairs*, *29*(4), 342–370.
Loeb, D. (1998). Moral realism and the argument from disagreement. *Philosophical Studies, 90*(3), 281–303.
Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Baltimore: Penguin.
McGrath, S. (2007). Moral disagreement and moral expertise. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 4, pp. 87–108). Oxford: Oxford University Press.
Moody-Adams, M. M. (1997). *Fieldwork in familiar places: Morality, culture, and philosophy*. Harvard University Press.
Murray, S. L. (1999). The quest for conviction: Motivated cognition in romantic relationships. *Psychological Inquiry, 10*(1), 23–34.
Nietzsche, F. W. (1886/1990). *Beyond good and evil: prelude to a philosophy of the future.* Baltimore: Penguin Books.
Parfit, D. (2011). *On What Matters*. Oxford University Press.
Plunkett, D., & Sundell, T. (2013). Disagreement and the semantics of normative and evaluative terms. *Philosophers' Imprint, 13*(23), 1–37.
Pritchard, D. (2008). Sensitivity, safety, and anti-luck epistemology. In J. Greco (Ed.), *The Oxford handbook of skepticism*. Oxford: Oxford University Press.
Ratchnevsky, P. (1993). *Genghis Khan: His life and legacy*. New York: Wiley.
Sauer, H. (2012). Educated intuitions. Automaticity and rationality in moral judgement. *Philosophical Explorations, 15*(3), 255–275.
Sayre-McCord, G. (1996). Coherentist epistemology and moral theory. In W. Sinnott-Armstrong & M. Timmons (Eds.), *Moral Knowledge? New Readings in Moral Epistemology*. Oxford: Oxford University Press.
Scanlon, T. M. (2014). *Being realistic about reasons*. Oxford: Oxford University Press.
Setiya, K. (2012). *Knowing right from wrong*. Oxford: Oxford University Press.
Shafer-Landau, R. (1994). Ethical disagreement, ethical objectivism and moral indeterminacy. *Philosophy and Phenomenological Research, 54*(2), 331–344.
Shafer-Landau, R. (2003). *Moral realism: A defence*. Oxford: Oxford University Press.
Smith, D. L. (2014). Dehumanization, essentialism, and moral psychology. *Philosophy Compass, 9*(11), 814–824.
Smyth, N. (2017). Moral knowledge and the genealogy of error. *Journal of Value Inquiry, 51*(3), 455–474.
Sripada, C., & Stich, S. (2006). A framework for the psychology of norms. In P. Carruthers, S. Laurence, & S. P. Stich (Eds.), *The innate mind, volume 2: Culture and cognition*. Oxford: Oxford University Press.
Street, S. (2009). In defense of future Tuesday indifference: Ideally coherent eccentrics and the contingency of what matters. *Philosophical Issues, 19*(1), 273–298.
Vavova, K. (2014). Moral disagreement and moral skepticism. *Philosophical Perspectives, 28*(1), 302–333.
Weatherson, B. (2010). Do judgments screen evidence? (unpublished manuscript).
Wedgwood, R. (2010). The moral evil demons. In R. Feldman & T. Warfield (Eds.), *Disagreement*. Oxford: Oxford University Press.
Wedgwood, R. (2014). Moral disagreement among philosophers. In M. Bergmann & P. Kain (Eds.), *Challenges to moral and religious belief: Disagreement and evolution* (pp. 23–39). Oxford: Oxford University Press.
Williams, B. (2009). The point of view of the universe: Sidgwick and the ambitions of ethics. In M. Burnyeat (Ed.), *The sense of the past: Essays in the history of philosophy* (pp. 277–296). Princeton: Princeton University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.