The Problem of Morally Repugnant Beliefs
Declan Smithies

What is the connection between justification and truth in moral epistemology? Epistemological orthodoxy says that knowledge is factive, while justification is non-factive: in other words, knowledge must be true, whereas justified beliefs can be false. This paper explores whether morality provides an exception to this general rule. Can you have justified false beliefs about your own moral obligations?

The prevailing view in moral epistemology is that you can. The sources of justification for moral belief are generally thought to include coherence (Sayre-McCord 1996), perception (McGrath 2004), intuition (Huemer 2005), emotion (Döring 2007), and testimony (Sliwa 2012). Moreover, all these sources of justification are widely thought to be non-factive. These assumptions make it all but inevitable that you can have justified false beliefs about your own moral obligations.

The primary goal of this paper is to argue that the prevailing view is false: you cannot have justified false beliefs about your own moral obligations. Otherwise, you could be justified in acting on your justified false beliefs about what you should do. If these actions are justified, however, then you need no excuse to avoid culpability for performing them. And yet surely you’re blameworthy for violating your moral obligations unless you have some excuse. That is my argument in a nutshell.

A secondary goal of the paper is to explain why you cannot have justified false beliefs about your own moral obligations. Some epistemologists explain this by embracing a global truth-connection in epistemology, which says that all epistemic justification is factive (Littlejohn 2012, Williamson 2019). Personally, I regard this as too radical a departure from epistemological orthodoxy. Instead, I endorse a local truth-connection in moral epistemology, which says that epistemic justification is factive when it concerns your own moral obligations. To explain this, we need a version of moral rationalism, which
says there are necessary truths about morality that everyone has a priori justification to believe. So the ultimate conclusion of the paper is that we have strong reasons to accept moral rationalism.

1. Introducing the Problem

What’s wrong with the prevailing view that we can have justified false beliefs about our own moral obligations? I’ll argue that it faces a serious problem, which I call the problem of morally repugnant beliefs. The problem, in short, is that it allows that people are sometimes justified in holding false beliefs that are morally repugnant. But this much is hard to countenance. I’ll argue that we should reject any moral epistemology on which morally repugnant beliefs can be epistemically justified.

Let’s begin by reviewing how this problem arises in the extant literature. Thomas Kelly and Sarah McGrath (2010) argue against a version of moral coherentism, which says that reflective equilibrium is sufficient for epistemic justification. Reflective equilibrium is the process of adjusting moral beliefs about particular cases in light of moral beliefs about general principles, and vice versa, until your moral beliefs cohere with each other and with your non-moral beliefs. According to moral coherentism, the resulting beliefs are epistemically justified (Sayre-McCord 1996). The basic problem is that reflective equilibrium is not discriminating enough to generate justified outputs from unjustified inputs. Hence, moral coherentism confronts a dilemma. If we demand that the inputs must be justified in some other way, then reflective equilibrium is not sufficient for epistemic justification. But if we impose no such demand, then nothing prevents the process from yielding unjustified outputs from unjustified inputs. Garbage in, garbage out.

To illustrate the point, Kelly and McGrath note that the inputs to the process might include belief in absurd principles of morality, for example:

One is morally required to occasionally kill randomly. (2010: 347)
According to John Rawls (1971), for example, the inputs to reflective equilibrium are “considered judgments” made confidently in conditions free from distraction, self-interest, and bias. In principle, however, there is no reason why someone in these conditions cannot make the considered judgment that this absurd moral principle is true. As Kelly and McGrath explain, “there is nothing incoherent about the possibility that someone could confidently and stably subscribe to this judgment, even if he or she is aware of all of the non-moral facts, does not stand to gain or lose depending on whether it is true or false, and so on” (2010: 347).

Moral coherentism therefore confronts the following objection: (1) If coherence is sufficient for epistemic justification, then an ideally coherent agent can be epistemically justified in believing it’s morally required to kill people randomly. (2) Intuitively, however, it’s never reasonable, and hence never epistemically justified, to hold this morally repugnant belief. Hence, (3) moral coherentism is false: coherence is not sufficient for epistemic justification.

The same problem arises for other prominent theories in moral epistemology. Consider Michael Huemer’s principle of phenomenal conservatism:

If it seems to S as if \( p \), then S thereby has at least prima facie justification for believing that \( p \).

(2001: 99)

On this view, seemings provide some degree of prima facie justification for believing their contents. The degree of justification provided by a seeming is proportional to its strength: the more strongly something seems true, the more justification you have to believe it. This prima facie justification can be defeated, but only by other seemings (Huemer 2001: 100; 2005: 100). When your seemings conflict,
what you’re justified in believing depends on the overall balance of how things seem. On this view, you’re epistemically justified in believing whatever seems strongly enough on balance to be true.

Clayton Littlejohn (2011) argues that phenomenal conservatism has the implausible consequence that terrorists can be epistemically justified in believing that they ought to kill people. We need only imagine that it seems strongly and unequivocally to some terrorist that he ought to kill people. Timothy Williamson makes the same point in the passage below:

Suppose that it seems to the consistent neo-Nazi that he ought to kill such people, with reference to some totally innocent members of one of the many groups neo-Nazis target, just because they belong to that group. Moreover, that intellectual seeming perfectly coheres with all his other seemings and beliefs, thanks to the harmonizing effects of his unconscious biases. As a result, he goes ahead and forms the belief that he ought to kill such people. By internalist standards, the neo-Nazi is justified in believing that he ought to kill such people . . . But is that really an appropriate way to view a consistent neo-Nazi? (2019: 163-4)

Phenomenal conservatism seems to imply that the ideally coherent Nazi is justified in believing that he ought to kill people just because of their demographic status. As Williamson protests, however, this verdict seems “outrageous”. I'll say more about this in due course, but to a first approximation, it threatens to undermine our condemnation of beliefs that we regard as morally beyond the pale.

The problem of morally repugnant beliefs has further applications that have not yet been explored in the literature. Consider the view that we can acquire moral knowledge based on sensory perception (McGrath 2004). When we see someone torturing a cat, for example, perhaps we can know by moral perception that they are doing something morally wrong. If so, then why can’t we undergo moral illusions in which we acquire justification to believe that torturing cats is morally permissible or
even required? A similar problem arises for the view that we can acquire moral knowledge based on emotions that represent values (Döring 2007). If we can know that torturing cats is immoral based on the negative emotion of moral disapproval, then why can’t someone with warped moral sensibilities acquire the justified belief that torturing cats is morally permissible based on the positive emotion of moral approval?

My primary goal here is to argue that the problem arises for any view that allows for justified false beliefs about your own moral obligations. To solve the problem, we must embrace a factive conception of epistemic justification for belief about your own moral obligations. But my argument for this conclusion will emerge from my criticisms of alternative strategies for solving the problem.

According to what I’ll call the divide-and-conquer strategy, we can solve the problem of morally repugnant beliefs by distinguishing between epistemic and moral dimensions of criticism. On this view, the Nazi is epistemically justified in holding beliefs that are nevertheless morally repugnant to have or to act upon. The hope is to avoid the appearance of condoning Nazi ideology by combining a positive epistemic evaluation with a negative moral evaluation.

I’ll outline three versions of this strategy, which diverge in their moral criticism of the Nazi. One relies on a distinction between epistemic and moral standards of justification for belief: it says the Nazi’s belief is epistemically justified, but morally unjustified. The second relies on a distinction between justification for belief and justification for action: it says that while the Nazi’s belief is epistemically and morally justified, he is not morally justified in acting on his justified belief. The third relies on a distinction between justification and value: it says that while the Nazi is morally justified in acting on his belief, he does something morally bad, although he does nothing morally wrong.

In the next three sections, I criticize each of these versions of the divide-and-conquer strategy. In the penultimate section, I combine these criticisms to give a deductive argument for the conclusion that the Nazi cannot be epistemically justified in holding his morally repugnant belief. Finally, I explain
this conclusion by appealing to a version of moral rationalism that says there are necessary truths about morality that everyone has a priori justification to believe. One such example is that it’s always wrong to commit genocide – that is, to kill people just because of their demographic status.

Let me begin by stating some assumptions up front. First, there are some moral facts. Second, there are moral facts about which actions are morally justified given our limited evidence as well as moral facts about which outcomes are morally best given all the facts. Third, these can come apart: we can be morally justified in performing actions whose outcomes we know to be morally suboptimal (Jackson 1990; Weatherson 2019: 28-30). To forestall confusion, I’ll use deontic terms (e.g. ‘justified’, ‘permitted’, ‘obligated’) for evidence-sensitive moral facts and evaluative terms (e.g. ‘good’, ‘best’) for evidence-insensitive moral facts. These assumptions are not question-begging, since we can suppose that our moral obligations are sensitive to our evidence without prejudging whether our evidence can justify false beliefs about our moral obligations.

2. Moral and Epistemic Justification for Belief

The first version of the divide-and-conquer strategy invokes a distinction between epistemic and moral standards of justification for belief. According to this strategy, beliefs that are justified by epistemic standards can be unjustified by moral standards. For example, the Nazi may be epistemically justified and yet morally unjustified in believing that he ought to commit genocide. As Geoffrey Sayre-McCord writes, “Recognizing that a person is epistemically justified in holding her beliefs may still leave us thinking the beliefs not just false but – as will often enough be the case with people’s moral views – repugnant” (1996: 148).

This strategy relies on two assumptions. The first assumption is that beliefs can be justified by practical and moral standards as well as epistemic standards. The second assumption is that these standards of justification can conflict. Thus, Simon Keller (2004) and Sarah Stroud (2006) argue that
the moral requirements of friendship can conflict with the epistemic requirement to respect your evidence. Suppose you acquire evidence that would normally justify believing that someone you know has let you down. If that person is your friend, however, the moral demands of friendship may require giving them the benefit of the doubt. Allegedly, the belief that your friend has let you down may be justified by epistemic standards but unjustified by moral standards.¹

I argue that we should accept the first assumption but not the second. We do sometimes evaluate beliefs by moral standards of justification. For instance, we can justly criticize people on moral grounds for failing to respect their evidence. Suppose that residual anger over some previous disagreement motivates you to jump too quickly to the conclusion that your friend has let you down. In that case, you wrong your friend by forming this belief without adequate evidence. Even if you never act on this belief, the mere fact that you hold it is enough to constitute a moral injustice. If she finds out, your friend can legitimately complain that your belief is unfair.

The situation is different, however, when your evidence is strong enough to provide epistemic justification for belief. Consider the “good case” in which you know that your friend has let you down. In that case, your friend has no legitimate complaint against you, since you’ve done nothing wrong. You cannot wrong someone just by knowing a fact about them.

Now consider the “bad case” in which you have the very same evidence, although your evidence is misleading. You can’t know your friend has let you down, since knowledge is factive, but your false belief is epistemically justified by your misleading evidence. Just as before, your friend has no legitimate complaint against you, since you’ve done nothing wrong. You cannot wrong someone just by proportioning your beliefs to your evidence. If your friend knows that your belief is false, she

¹ Harman (2011: 462) mentions these cases to support the possibility that false moral beliefs might be epistemically justified but morally blameworthy. She claims that we have moral obligations to believe moral truths and to avoid believing moral falsehoods.
may try to persuade you by providing new evidence. Given this new evidence, you may realize that your belief was false and come to regret holding it. And yet you did nothing wrong merely by following your evidence where it leads. Your false belief may be regrettable, in retrospect, but it isn’t wrong.

On some views, there is moral encroachment in the standards for epistemic justification. Perhaps friendship raises the threshold for how much evidence you need to be epistemically justified in believing that your friend has let you down. If so, then your evidence may justify high confidence that your friend has let you down without being strong enough to justify outright belief. In that case, the moral demands of friendship may require giving your friend the benefit of the doubt. But this view doesn’t imply that the standards for epistemic justification can conflict with the standards of morality.² If your evidence is strong enough to make your belief epistemically justified given the high moral stakes, then it cannot plausibly be morally unjustified at the same time.

To sum up, morality may sometimes require respecting your evidence, but it never requires disrespecting your evidence or abandoning your knowledge. This means there can be no conflict between epistemic and moral standards of justification for belief:

*No Conflicts:* If your belief is justified by epistemic standards, then it is justified by moral standards, and justified simpliciter.

This can be further explained as a consequence of evidentialism, the thesis that all justifying reasons for belief are evidential reasons. Evidentialism, so defined, is consistent with moderate forms of moral

---

² Gendler (2011) argues that moral and epistemic requirements sometimes conflict because avoiding racist beliefs requires neglecting base rates, as in her example of John Hope Franklin at the Cosmos Club. In reply, Basu and Schroeder (2019) argue for moral encroachment on the grounds that it avoids such conflicts, while Gardiner (2019) argues that such conflicts can be avoided without commitment to moral encroachment.
and pragmatic encroachment in epistemology, according to which non-evidential factors may determine how much evidence is needed to justify outright belief. What it excludes is the possibility that non-evidential reasons for belief, including moral reasons, can conflict with evidential reasons. According to evidentialism, there are no non-evidential reasons for belief.

Some argue that evidentialism faces counterexamples. Suppose, for example, that an evil demon threatens to destroy the world unless you believe an explicit contradiction. Don’t you thereby have moral reasons for belief that justify disrespecting your evidence? Arguably, not. Your moral reasons justify acting in ways designed to bring it about that you believe a contradiction, but they don’t thereby justify believing a contradiction. We should reject the general principle that anything that gives you a justifying reason to bring it about that you φ thereby gives you a justifying reason to φ. Suppose an evil demon threatens to destroy the world unless you have a headache. This gives you a reason to bring it about that you have a headache, but it gives you no reason for the headache itself. After all, a headache is not the kind of state for which you can have justifying reasons.

A justifying reason to φ is often glossed as “a consideration that counts in favor” of φ-ing (Scanlon 1998: 17). Admittedly, there is some sense in which the demon’s threat counts in favor of having a headache, or believing a contradiction, but not in the right way to count as a justifying reason. In this context, a justifying reason to φ is a consideration that can make it reasonable, rational, or justified to φ. The demon’s threat is a consideration in light of which it’s good to believe contradictions, but it’s not a consideration that makes it reasonable to believe contradictions. Rather, it’s a consideration that makes it reasonable to act in ways designed to bring it about that you believe a contradiction.

This conception of justifying reasons suggests the following argument for evidentialism:

(1) All justifying reasons to believe that p can make it reasonable to believe that p.

(2) Only evidential reasons to believe that p make it reasonable to believe that p.
Therefore, all justifying reasons to believe that \( p \) are evidential reasons to believe that \( p \).³

The first premise is supported by the conception of justifying reasons for belief as considerations that count in favor of belief by making them reasonable, rational, or justified. The second premise is supported by the consideration that it’s unreasonable to base your beliefs on moral or practical considerations alone. For example, it’s unreasonable to infer a contradiction from the premise that believing this conclusion will save the world from destruction. This is fallacious reasoning motivated by wishful thinking. No doubt, it would be good to engage in such reasoning when this would placate the demon. Anyone who knows this has decisive moral reasons to do whatever they can to induce such reasoning. And yet this doesn’t make it reasonable to engage in fallacious reasoning. Rather, it makes it reasonable to act in ways designed to make your reasoning unreasonable (Parfit 1984: 12-13).

Evidentialism is not beyond dispute, but it remains plausible enough to be regarded as a default position in epistemology. It would be surprising if we cannot solve the problem of morally repugnant beliefs without abandoning evidentialism. In what follows, I’ll consider various evidentialist solutions to the problem. We can criticize racist beliefs by moral standards when they conflict with the evidence in ways that are motivated by immoral attitudes, such as racial hatred. And yet we cannot extend the same criticism to Williamson’s ideally coherent Nazi on the assumption that his beliefs are epistemically justified. We must therefore look elsewhere for an evidentialist solution to the problem of morally repugnant beliefs.

---

³ Compare Way’s (2016) argument for evidentialism, which relies on the premise that reasons are premises in good reasoning. My argument diverges in two ways: first, I explicate good reasoning in terms of reasonableness; and second, I don’t assume all reasonable beliefs are formed by reasoning.
3. Justification for Belief and Action

The second version of the divide-and-conquer strategy relies on a distinction between justification for belief and justification for action. Suppose the Nazi’s belief is justified not only by epistemic standards, but also by moral standards, and justified simpliciter. Are we thereby committed to saying that the Nazi is justified in acting on his belief? This strategy says no: the Nazi is justified in believing that he ought to commit genocide, but he is not justified in committing genocide.

Again, this strategy relies on two assumptions. The first assumption is that belief and action are justified by different standards: belief is justified by epistemic standards, whereas action is justified by practical standards, including morality. The second assumption is that epistemic and practical standards can diverge in such a way that you’re justified in holding beliefs that you’re not justified in acting upon. I argue that we should accept the first assumption but reject the second.

Williamson (2019) also rejects the second assumption on the grounds that it drives too thick a wedge between belief and action. His central claim is that the epistemic norms for belief reflect its functional role in guiding action. Thus, he writes:

The nature of belief is that the believer is disposed to act on it . . . Given that connection with action in the nature of belief, we can expect it to be reflected in norms for belief. (2019: 167)

Personally, I don’t accept Williamson’s functionalist thesis that the nature of belief is to guide action. Even so, we can replace this with the normative thesis that practical rationality requires agents to be guided by their beliefs in action. Moreover, we can expect this normative connection between belief and action to be reflected in the epistemic norms for belief. Unfortunately, Williamson doesn’t spell out this normative connection between belief and action in detail and so his argument remains enthymemetic. In this section, I’ll flesh out the argument in my own way.
We can begin with the general principle that if you’re justified in believing something, then you’re justified in treating it as a reason for action. As Fantl and McGrath articulate the point, anything you’re justified in believing is warranted enough to serve as a justifying reason for action:

If you are justified in believing that \( p \), then \( p \) is warranted enough to justify you in \( \phi \)-ing, for any \( \phi \). (2009: 99)

This cannot plausibly mean that any action is justified by anything else you’re justified in believing. You’re not justified in sunbathing, for example, just because you’re justified in believing that you’re dehydrated. After all, you’re not justified in believing that dehydration is a reason to sunbathe. Presumably, though, you’re justified in believing that dehydration is a reason to drink water. That’s why your justification to believe that you’re dehydrated also justifies drinking water. This is just one instance of a more general normative connection between belief and action:

*The Belief-Action Link:* If you’re justified in believing that \( p \), and you’re justified in believing that \( p \) is a sufficient reason to \( \phi \), then you’re justified in \( \phi \)-ing for the reason that \( p \).

Now we can apply this general principle to the special case of beliefs about moral obligation. I assume that we’re always justified in believing that our moral obligations give us sufficient reasons for action. So, if the Nazi is justified in believing that he morally ought to commit genocide, then he is justified in committing genocide for the reason that he morally ought to do so. And yet this is a bitter pill to

---

4 Here, and throughout, I use ‘justified’ in its ex ante or forward-looking sense.
swallow. As Williamson writes, “Is it not simply outrageous . . . to claim that the consistent neo-Nazi is justified in killing such people?” (2019: 170)

Williamson (2019: 172, n. 7) avoids this conclusion by claiming that only known facts can be justifying reasons for action. On this view, you’re justified in acting for the reason that p only if you know that p. Hawthorne and Stanley articulate this connection as follows:

*The Reasons-Knowledge Principle*: When one’s choice is p-dependent, it is appropriate to treat the proposition that p as a reason for acting iff you know that p. (2008: 578)

Of course, the Nazi cannot *know* that he ought to commit genocide, since knowledge is factive. Therefore, Williamson concludes, he is not justified in committing genocide for the reason that he ought to do so. This blocks the outrageous conclusion.

Given the Belief-Action link, Williamson must deny that the Nazi is justified in holding the false belief that he ought to commit genocide. He therefore rejects the orthodox view that epistemic justification is non-factive. Instead, he endorses a knowledge norm for belief, which says that “a belief is fully justified if and only if it constitutes knowledge” (2013: 5). I’ll revisit this solution in §5, but we can set it aside here, since the strategy we’re now considering is that the Nazi has a justified false belief about his moral obligations.

So we can assume for now that epistemic justification is non-factive: you can be justified in believing false propositions on the basis of misleading evidence. The Belief-Action Link implies that you’re justified in treating these false propositions as reasons for action. Moreover, this is supported by reflection on examples. Suppose you’re planning to meet a friend, but she texts you the wrong location by mistake. Plausibly, your misleading evidence can justify not only believing that your friend is on Main St, but also taking the bus to Main St in order to meet your friend. Your action is regrettable
in retrospect, since your friend isn’t there. Nevertheless, you’re justified in treating the false consideration that your friend is on Main St as a reason for taking the bus there. If so, then why isn’t the Nazi justified in treating the false consideration that he ought to commit genocide as a reason to commit genocide?

One response is that the Belief-Action Link doesn’t apply in this example because normative facts cannot provide reasons for action. Some proponents of “buck-passing” argue that moral obligations cannot give you reasons for action, since reasons for action are already provided by non-moral facts that ground those moral obligations. If we suppose that your moral obligations give you additional reason for action, then we engage in an objectionable kind of double-counting. In reply, however, the consideration that you have a moral obligation to act does seem to “count in favor” of action in the right way to count as a justifying reason for action. And the double-counting objection can be blunted by noting that the reasons provided by normative facts depend on the reasons provided their non-normative grounds in much the same way that the normative facts themselves depend on their non-normative grounds (Johnson King 2019).

To block the outrageous conclusion, friends of the divide-and-conquer strategy must give independent reasons to reject the Belief-Action Link. One objection is that high stakes cases generate counterexamples in which you’re not justified in acting upon your justified beliefs. Consider the high stakes version of Keith DeRose’s (1992) bank case in which you have evidence that the bank will be open tomorrow, but the costs of being wrong are so high that you’re not justified in coming back tomorrow, rather than waiting in line today. Even so, the objection continues, your evidence might be strong enough to justify believing that the bank will be open tomorrow.

A hardline response is that you’re not justified in believing the bank will be open, rather than merely having a high degree of confidence, unless you’re justified in treating what you believe as a sufficient reason for action (Fantl & McGrath 2009). On this view, the stakes affect how much
evidence you need to be justified in holding the belief, since it is justified when the stakes are low, but not when the stakes are high. The response is controversial, since it endorses pragmatic encroachment in epistemology. Fortunately, a more concessive response is available. Suppose you’re justified in believing that the bank will be open, but given the high stakes, you’re unjustified in treating this as a sufficient reason to come back tomorrow. This doesn’t undermine the Belief-Action Link unless you’re also justified in believing that the consideration that the bank will be open gives you a sufficient reason to come back tomorrow. And yet we cannot plausibly combine all these assumptions.

To see why, suppose you’re justified in believing not only (i) that the bank is open tomorrow, but also (ii) that if the bank is open tomorrow, then you should come back tomorrow. Assuming that justification is closed under entailment, you’re thereby justified in believing (iii) that you should come back tomorrow. Given these assumptions, that is what you should believe you should do. But we’re also supposing that this conclusion is false: given the high stakes, you should wait in line, rather than coming back tomorrow. So we get the implausible result that you should do the exact opposite of what you should believe you should do. And yet akrasia – that is, acting against your beliefs about what you should do – is a paradigmatic form of irrationality. Plausibly, you should never be akratic.

More generally, rejecting the Belief-Action Link has the implausible result that sometimes you may – and, indeed, should – be akratic. Suppose the Nazi is justified in believing that he has a moral obligation to commit genocide, and that he is justified in believing that this moral obligation gives him a sufficient – indeed, overriding – reason to commit genocide. In that case, he should believe that he should commit genocide for the reason that it’s his moral duty. But now suppose that he shouldn’t commit genocide because it is unjustified and hence impermissible. This implies that the Nazi should
be akratic – that is, he should refrain from doing something that he should believe he should do. And yet this is highly implausible. Again, you should never be akratic.⁵

Nomy Arpaly (2003) argues that there are cases of inverse akrasia in which it is rationally and morally better to act in conflict with your moral beliefs than to act in accordance with them. Thus, Huckleberry Finn should help Jim to escape slavery, rather than turning him over to the authorities, despite his own belief to the contrary. However, this doesn’t imply that he may or should be akratic. After all, he should believe that he should help Jim, rather than believing that he should turn him in. Just as Huck’s recognition of Jim’s humanity provides good moral reasons to help him, so it provides him with good epistemic reasons to believe that he should help him.⁶ Although his action is rationally responsive to these good reasons, his belief is not rationally responsive to them, and so there is some irrationality in his overall akratic predicament. As Arpaly herself notes, “every agent who acts against her best judgment is, as an agent, less than perfectly rational, as the schism between best judgment and desire indicates a failure of coherence in her mind” (2003: 36).

Some philosophers maintain that akrasia is justified when you’re justified in holding false beliefs about what you should do.⁷ On this view, you might be justified in believing that you should act for some reason without being justified in acting for that reason. In such cases, you should refrain from acting for the reasons you should think you should act on. Nevertheless, this view is unattractive. After all, akrasia is a paradigmatic form of irrationality. Rationality requires coherence – in particular,

---

⁵ The argument assumes that justification has deontic consequences: you should believe what you’re justified in believing and you shouldn’t do what you’re not justified in doing. Indeed, since you may never be akratic, the argument goes through on the weaker assumption that you may believe what you’re justified in believing and you may not do what you’re not justified in doing.

⁶ Harman (2022) argues that Huck Finn is praiseworthy for his actions, but blameworthy for his beliefs, because only his actions are responsive to the reasons provided by his recognition of Jim’s humanity. In contrast, I claim only that his beliefs are unjustified, which doesn’t imply that they are blameworthy.

⁷ Weatherson (2019: Ch. 10) and Lasonen-Aarnio (2020) defend “level-splitting views” on which akrasia is sometimes justified. Meanwhile, Titelbaum (2015) and Smithies (2019: Chapters 9 & 10) defend the opposing view that akrasia is never justified.
it requires that your actions cohere with your beliefs about how you should act. As such, akrasia always involves some degree of irrationality. And nothing can justify being irrational.\(^8\)

Instead, we should turn this argument on its head. Since akrasia is never justified, you cannot be justified in holding false beliefs about what you should do. We should therefore accept what Clayton Littlejohn (2011: 41) calls ‘the JJ Principle’:

\[
\text{The JJ Principle: If the agent justifiably judges she should } \phi, \text{ the agent is justified in } \phi\text{-ing.}\quad \text{9}
\]

The JJ Principle is just an instance of the Belief-Action Link applied to the special case of normative beliefs about what you should do: if you’re justified in believing that you should \(\phi\), then you’re justified in \(\phi\)-ing for the reason that you should do so. Anyone who rejects the JJ Principle is committed to the implausible consequence that akrasia can be justified. If we accept the JJ Principle, however, then we must choose between two options. One option is \textit{modus ponens}: the Nazi is justified in committing genocide, since he is justified in believing he ought to do so. The other option is \textit{modus tollens}: the Nazi is not justified in believing he ought to commit genocide, since he is not justified in acting on this belief. I argue against the first option in §4 and defend the second in §5.

\section*{4. Justification and Value}

The third version of the divide-and-conquer strategy draws a distinction between \textit{justification} and \textit{value}.

On this view, the Nazi is not only justified in believing he ought to commit genocide, but he is also

---

\(^8\) Worsnip (2018) argues that the structural requirement to be rational can conflict with the substantive requirement to respond to reasons. On this bifurcationist view, you sometimes have substantive reasons to be structurally irrational. In contrast, I assume the unified view that your reasons can never justify being irrational. See Smithies (forthcoming a) for further discussion.

\(^9\) Whiting and Way defend the stronger principle of Ought Infallibilism: “If you justifiably believe that you ought to \(\phi\), you ought to \(\phi\)” (2016: 1873).
justified in acting on this belief. Even so, we can criticize his beliefs and actions on the grounds that they have morally bad consequences. On this view, the Nazi violates no moral obligation, since he is morally justified in believing and acting as he does. Nevertheless, in having and acting on these beliefs, the Nazi does something morally bad, although he does nothing morally wrong.

This strategy relies on two assumptions. The first assumption is that people can be justified in doing things that have morally bad consequences. The second assumption is that this explains what’s wrong with the Nazi’s morally repugnant beliefs and actions. Once again, I accept the first assumption and reject the second.

In support of the first assumption, consider the Unwitting Doctor who prescribes some medicine in order to cure a patient’s illness, although it has an unforeseen side-effect that kills him. Assuming that epistemic justification is non-factive, let’s suppose she has misleading evidence that justifies believing that the medicine will cure her patient and hence that she ought to prescribe it. Given the Belief-Action Link, it follows that she is justified in prescribing the medicine for the reason that it will cure him. Since she is justified in prescribing the medication, it cannot be wrong for her to do so. By inadvertently killing the patient, she does something tragic, and regrettable, but she does nothing wrong. Indeed, it would be wrong for her to withhold the medicine, since she has every reason to believe it will cure him. Of course, she should regret killing the patient, and apologize to his family, but that doesn’t mean her action is morally unjustified.\(^\text{10}\)

The second assumption is that the Nazi’s moral ignorance justifies his action in the same way as the Unwitting Doctor’s non-moral ignorance. Intuitively, however, there is a crucial asymmetry between moral and non-moral ignorance. The Unwitting Doctor does something bad, but she does nothing wrong. In contrast, the Nazi does something wrong and not merely something bad. In my

\(^{10}\) Williams (1976) argues that a truck driver who kills a child through no fault of his own should regret his involvement, and apologize to the child’s family, although he does nothing morally wrong.
view, any plausible solution to the problem of morally repugnant beliefs must explain the asymmetry between these cases. The main goal of this section is to motivate this *asymmetry constraint*.

We can motivate this asymmetry constraint by appealing to considerations about culpability. It seems inappropriate to blame the Unwitting Doctor for prescribing the medicine when her evidence justifies believing that it will cure her patient. In contrast, the Nazi’s genocidal actions deserve moral condemnation. As Williamson insists, we should blame the Nazi for what he does:

If internalists take consistency with the agent’s perspective to bring blamelessness with it, then in evaluating the neo-Nazi’s beliefs and actions as justified, because consistent with his perspective, they are implying that his beliefs and actions are blameless. But the neo-Nazi *should* be blamed for killing those innocent people. (2019: 171)

We compromise our moral condemnation of the Nazi if we concede that he is justified in committing genocide, since it’s inappropriate to blame someone for justified action. We should therefore deny that the Nazi is justified in committing genocide.

Call this the Moral Culpability Argument:

(1) *The Culpability Datum*: The Nazi is culpable for committing genocide.

(2) *The Justification-Culpability Link*: If you’re justified in φ-ing, then you’re not culpable for φ-ing.

(3) *Conclusion*: Therefore, the Nazi is not justified in committing genocide.

---

11 One exception we can set aside is that justified actions may be culpable in a derivative way when this derives from culpability for some other unjustified action. For instance, you may be derivatively culpable for acting on beliefs that are justified only because of culpably negligent methods of gathering evidence in the past.
The weakest link in this argument is the Culpability Datum. In the literature on moral responsibility, there is some dispute about whether the Nazi is culpable for his actions. Perhaps he is someone who must be stopped, but not someone who can be blamed. Violating a moral obligation is not sufficient for culpability, since you’re not culpable for wrongdoing when you have an excuse. Moreover, one such excuse is that you cannot reasonably be expected to know that you’re doing anything wrong. And we can imagine that the Nazi has such an excuse – say, because he is brainwashed by a morally corrupt community. If so, then perhaps the Nazi is not culpable for his actions.

Gideon Rosen gives a similar defense of an ancient slaveholder who “buys and sells human beings, forces labor without compensation, and separates families to suit his purposes” (2003: 64). Although he acts wrongfully, he believes otherwise – not because he thinks slaves have any innate psychological or biological inferiority, but merely because he has internalized the normative standards of his society. So the slaveholder acts from moral ignorance. But Rosen endorses the following thesis:

_The Parity Thesis:_ Whenever an agent acts from ignorance, whether factual or moral, he is culpable for the act only if he is culpable for the ignorance from which he acts. (2003: 64)

And he argues that the slaveholder is not culpable for his moral ignorance because we cannot reasonably expect him to know any better: “Given the intellectual and cultural resources available to a second millennium Hittite lord, it would have taken a moral genius to see through to the wrongness of chattel slavery” (2003: 66). Hence, Rosen concludes, he is not culpable for his wrongful actions.

In response, Elizabeth Harman (2011, 2022) argues that moral ignorance cannot exculpate wrongful action in the same way that non-moral ignorance can. On her view, the ancient slaveholder is blameworthy for his actions even if he fulfils all his procedural obligations in the management of his beliefs. She accepts Rosen’s thesis that he is blameworthy for his action only if he is blameworthy
for the ignorance from which he acts. But she argues that the slaveholder is blameworthy not only for his actions, but also for his moral beliefs, and for much the same reason – namely, that they manifest his failure to care adequately about what is morally important.

My own sympathies lie with Rosen’s position. At a minimum, the difficulty of acquiring moral knowledge is an excuse that can mitigate blame without removing blame entirely (Sliwa 2020). This is why we blame contemporary slaveowners more harshly than ancient slaveowners given the evolution in societal attitudes towards slavery. But difficulty comes in degrees: the harder it is to acquire moral knowledge, the better your excuse, and the less blameworthy you are for wrongdoing. Arguably, a high enough degree of difficulty can provide full exculpation for unjustified action. In particular, it seems inappropriate to blame someone for wrongdoing when it’s so difficult to acquire moral knowledge in the circumstances that it’s practically impossible. I’m certainly not claiming that this was true in Nazi Germany, but abstracting away from the contingencies of human history, we may suppose that our hypothetical Nazi satisfies this condition – say, because he is brainwashed by his community.¹²

Harman’s view is that culpability depends on your quality of will. On this view, the Unwitting Doctor is blameless because she acts from good will, whereas the Nazi is blameworthy insofar as he acts from ill will or moral indifference. But we can imagine a hypothetical Nazi who is motivated not by racial hatred, or callous indifference to human life, but rather by misdirected moral concern. He wants to do the right thing, but his corrupt moral education leads him astray, and he acts on his mistaken moral conclusions. It’s true that he isn’t motivated by what matters morally, rather than what he falsely believes to matter morally, but the question is whether he is culpable for being motivated

¹² Compare Alston’s (1985: 67-8) suggestion that “cultural isolation” can serve as an exculpation for unjustified beliefs. Given Rosen’s parity thesis, the same exculpation extends to any unjustified actions based on those unjustified beliefs.
by the wrong things. If we stipulate that his circumstances make it practically impossible for him to do any better, then it seems to me unreasonable to blame him.

In any case, we can revise the Moral Culpability Argument without taking any stand in this debate between Harman and Rosen:

(1) The Revised Culpability Datum: The Nazi is culpable for committing genocide unless he has an excuse.

(2) The Revised Justification-Culpability Link: If you’re justified in φ-ing, then you need no excuse to avoid culpability for φ-ing.

(3) Conclusion: Therefore, the Nazi is not justified in committing genocide.

Both sides can agree that the Nazi needs an excuse to avoid blame for committing genocide, while disagreeing about whether any such excuse is available. If the Nazi is justified in committing genocide, however, then he has no need for an excuse. An excuse is needed to absolve you from blame only when you violate your obligations. But when your actions are justified, you violate no obligations. Therefore, we should deny that the Nazi is justified in committing genocide.

As we’ve seen, Harman defends a version of the asymmetry constraint: she argues that non-moral ignorance can exculpate wrongdoing in a way that moral ignorance cannot. On this view, the Unwitting Doctor is not blameworthy for killing her patient because she has an excuse for wrongdoing, whereas the Nazi is blameworthy because he has no such excuse. My own position diverges from Harman’s in two respects. First, I deny that the Unwitting Doctor needs an excuse to avoid blame for killing her patient, since her justified non-moral ignorance justifies her action: she
does something bad, but she does nothing wrong. In contrast, the Nazi needs an excuse to avoid blame for committing genocide because his moral ignorance doesn’t justify his action: he does something wrong and not merely something bad. Second, however, the Nazi may be excused from blame for wrongdoing if his circumstances make it sufficiently difficult or even practically impossible to acquire the moral knowledge that he is doing anything wrong.

My own version of the asymmetry constraint concerns justification, rather than culpability. I contend that non-moral ignorance can justify action in a way that obviates any need for exculpation, whereas pure moral ignorance never justifies action, although it sometimes provides an exculpation. The problem with the third divide-and-conquer strategy is that it violates this asymmetry constraint. Perhaps the Nazi can be excused for violating a moral obligation, but we cannot plausibly say that he violates no moral obligation at all.

My argument here is Moorean in spirit. Assuming there are any moral facts at all, this is surely one of them: it’s always wrong to commit genocide – that is, to kill people just because of their demographic status. If some moral epistemology implies otherwise, then so much the worse for that theory. If we say instead that the Nazi is justified in committing genocide, then we capitulate to a version of the relativistic view that anything goes: there is no action, no matter how morally repugnant, that cannot be justified by appropriate evidence. To avoid this kind of overly permissive relativism, we must acknowledge that there are universal moral facts that hold necessarily for everyone, no matter what evidence they have. But if there are any such universal moral facts, I’ll argue, then we always have a priori justification to believe them.

13 Perhaps this disagreement is merely verbal, since Harman claims that the doctor’s action is wrong objectively, but not subjectively, whereas the slaveowner’s action is wrong in both senses. In any case, Harman’s theory of blameworthiness is not required for maintaining any such normative asymmetry.
5. Solving the Problem

We’ve now considered three versions of the divide-and-conquer strategy and found problems with them all. By combining the main points from our discussion, we can construct a deductive argument that the Nazi cannot be epistemically justified in believing that he ought to commit genocide:

(1) *No Conflicts*: If you’re epistemically justified in believing you ought to $\phi$, then you’re justified simpliciter in believing you ought to $\phi$.

(2) *The JJ Principle*: If you’re justified in believing you ought to $\phi$, then you’re justified in $\phi$-ing.

(3) *The Revised Justification-Culpability Link*: If you’re justified in $\phi$-ing, then you need no excuse to avoid culpability for $\phi$-ing.

(4) *The Revised Culpability Datum*: The Nazi is culpable for committing genocide unless he has an excuse.

(5) *Conclusion*: Therefore, the Nazi is not epistemically justified in believing he ought to commit genocide.

Moreover, the argument generalizes. Suppose, for *reductio*, that you’re epistemically justified in holding the false belief that you should $\phi$ when in fact you shouldn’t. By No Conflict, you’re justified simpliciter in believing that you should $\phi$. By the JJ Principle, you’re justified in $\phi$-ing for the reason that you should $\phi$. By the Revised Justification-Culpability Link, you need no excuse to avoid culpability for $\phi$-ing. In general, however, you need some excuse to avoid culpability for violating your moral obligations. Therefore, we must reject our initial assumption that your false belief about what you should do is epistemically justified.
This argument generates an explanatory challenge. Why can’t the Nazi have misleading evidence that epistemically justifies the false belief that he ought to commit genocide? More generally, why can’t you have misleading evidence that epistemically justifies false beliefs about what you should do? As we saw in §1, the prevailing view in moral epistemology cannot explain this because it assumes that the sources of epistemic justification for beliefs about your moral obligations are non-factive. In this concluding section, I consider two different strategies for answering the explanatory challenge.

One option is to invoke a general truth-connection in epistemology. Thus, Littlejohn (2012, 2014) and Williamson (2000, 2013) reject the orthodox view that justified beliefs can be false. On this view, epistemic justification is factive: you have justification to believe that \( p \) only if it’s a fact that \( p \). Indeed, Williamson’s (2000: 255-6) knowledge norm for belief implies that you have justification to believe that \( p \) only if you know that \( p \). As Williamson writes, “a belief is fully justified if and only if it constitutes knowledge” (2013: 5).

In my view, this is too radical a departure from epistemological orthodoxy. It is hard to deny that false beliefs can ever be justified by misleading evidence. For example, the Unwitting Doctor has misleading evidence that justifies the false belief that prescribing the medicine will cure her patient. Given the Belief-Action Link, it follows that she is justified in prescribing the medicine for the reason that it will cure him. Interestingly, Littlejohn (2012) argues in the opposite direction. He starts from the premise that the doctor is unjustified in prescribing the medicine because she violates a moral duty not to kill her patients. He therefore concludes that she is unjustified in her false belief that the medicine will cure him.

My main reservation about this solution is that it violates the asymmetry constraint. In §4, I argued that there is a normative asymmetry between moral and non-moral ignorance. For example, the Unwitting Doctor does something bad, although she does nothing wrong: she needs no excuse to avoid culpability, since she is justified in acting on her justified non-moral ignorance. In contrast, the
Nazi does something wrong and not merely something bad: he needs an excuse to avoid culpability because his moral ignorance doesn’t justify his action. The third version of the divide-and-conquer strategy violates the asymmetry constraint, since it implies that both the Nazi and the Unwitting Doctor are justified in their beliefs and actions. The current proposal violates the asymmetry constraint in the opposite direction, since it implies that neither are justified in their beliefs or actions.

Can we solve the problem in a way that respects the asymmetry constraint? My own solution is to endorse a *local truth-connection* in moral epistemology. More specifically, I endorse a version of moral rationalism:

*Moral Rationalism:* There are necessary truths about morality that everyone has sufficient a priori justification to believe.

For instance, everyone has sufficient a priori justification to believe the necessary moral truth that it’s always wrong to commit genocide. It’s not always wrong to kill people – say, in self-defense – but it’s always wrong to kill people just because of their demographic status. Since we always have sufficient a priori justification to believe this necessary moral truth, we never have sufficient justification to believe its negation, since we never have sufficient justification for contradictory beliefs. Hence, the Nazi cannot be justified in believing that he ought to commit genocide.

This explains why the Nazi’s moral ignorance doesn’t justify his action. He is not justified in committing genocide because he is justified in believing that doing so is immoral. That’s why he is blameworthy in the absence of an adequate excuse. Perhaps the Nazi can be excused from blame on the grounds that he cannot reasonably be expected to know he is doing anything wrong. His circumstances may even make it psychologically impossible for him to know the moral facts that he has a sufficient priori justification to believe. After all, there is no guarantee that a priori knowledge is
always psychologically feasible either in morality or elsewhere. Just as it’s unfair to blame someone for their actions when they cannot reasonably be expected to do otherwise, so it’s unfair to blame someone for their beliefs when they cannot reasonably be expected to know better. In such cases, you’re not blameworthy for your beliefs, but blamelessness is not sufficient for justification. Difficulty serves as an exculpation or an excuse, rather than a justification.

In contrast, the Unwitting Doctor’s justified non-moral ignorance justifies her action. After all, she has no evidence that justifies believing that the medicine will kill her patient. On the contrary, her evidence justifies believing that the medicine will cure him. That is why she is justified in giving him the medicine in order to cure him. Since her non-moral ignorance justifies her action, she needs no excuse to avoid culpability for the fatal consequences of her action.

More generally, when false non-moral beliefs are justified, they also justify action. This explains why we cannot generate counterexamples to the JJ Principle in which misleading evidence about your non-moral circumstances justifies false beliefs about your moral obligations. If your false non-moral beliefs are justified, then you’re also justified in acting on them. That is why we cannot use the problem of morally repugnant beliefs to argue that all epistemic justification is factive. What you’re morally justified in doing depends on what you’re epistemically justified in believing about your non-moral circumstances regardless of whether those beliefs are true or false.¹⁴

Of course, you can have misleading evidence about other people that justifies false beliefs about their moral obligations. After all, there is nothing inherently irrational about one person acting in conflict with another person’s beliefs about what they should do. In contrast, there is something inherently irrational about acting in conflict with your own beliefs about what you should do. This is

¹⁴ Given that epistemic akrasia is always unjustified, we cannot have epistemic justification for false beliefs about what we have epistemic justification to believe (Smithies 2019: Ch. 10).
one dimension of the asymmetry between first-person and third-person perspectives: you can have justified false beliefs about other people’s moral obligations, but not your own.

6. Challenges for Moral Rationalism

In conclusion, moral rationalism solves the problem of morally repugnant beliefs in a way that respects the asymmetry constraint. Even so, moral rationalism faces many challenges.

One such challenge is to delimit the scope of moral rationalism. Which are the necessary truths about morality that everyone has a priori justification to believe? I’ve endorsed the existential claim that there are some necessary truths about morality that everyone has a priori justification to believe: for example, it’s always wrong to commit genocide. I’m not thereby committed to the universal claim that everyone has a priori justification to believe every necessary truth about morality. Even so, any restrictions must be motivated in a principled way.

A second challenge is to explain moral rationalism. What guarantees that everyone has a priori justification to believe these necessary truths about morality? I’m not assuming that there is some a priori evidence that everyone is guaranteed to possess. In my view, that is the wrong model. Instead, the structure of the evidential support relation guarantees that there are necessary truths that are maximally supported by your evidence no matter what evidence you have. In a probabilistic framework, these necessary truths always have evidential probability one. Arguably, these include not only logical truths (Smithies 2015), and necessary truths about evidential support (Smithies 2019), but also necessary truths about morality (Smithies forthcoming b).

A third challenge is to defend moral rationalism against objections. Why can’t false testimony give you misleading evidence about necessary truths of morality? And why can’t your evidence make these necessary truths uncertain? My response is that you cannot have misleading or uncertain evidence about necessary truths that always have probability one. Of course, you may receive false
testimony from apparently reliable sources, but this doesn’t change the evidential probability of these necessary truths. This means you don’t always have evidential justification to believe the deliverances of false but apparently reliable testimony. There may be some non-ideal sense in which it’s appropriate to trust false testimony in such cases, but this isn’t sufficient for your beliefs to be evidentially justified (Smithies 2015, 2022).

It goes beyond the scope of this paper to address these challenges in any detail, but I pursue that project elsewhere. The main goal of this paper is not to defend moral rationalism against objections, but rather to explain why it is worth defending in the first place.15

---

15 I first presented this paper in my moral epistemology seminar in Fall 2019 and then at the Madison Meta-Ethics Workshop, the Cogito Epistemology Workshop, and the University of Pennsylvania in Fall 2021. I’m grateful to those audiences for helpful discussion and especially to Paul Boghossian, Jamie Fritz, Tristram McPherson, and anonymous referees for written comments.
References


