Moral Uncertainty and Moral Culpability

JAY GEYER
University of Colorado Boulder

Most of the literature on moral uncertainty has been oriented around the project of giving a normative theory for actions under moral uncertainty. The need for such a theory presupposes that internalist factors such as moral beliefs and evidence are relevant to what an agent ought to do. Some authors, including Elizabeth Harman, reject that presupposition. Harman advances an argument against all such internalist views on the grounds that they entail the exculpation of agents who should strike us as morally culpable. I argue that Harman's argument is only sound with respect to a small subset of internalist views, a subset that no one in fact defends. Though Harman's argument misses its mark, it raises important questions about how internalist theories should be understood. I argue that internalist theories should be understood as issuing rational, not moral, prescriptions.

I. INTRODUCTION

The moral uncertainty literature to date has been largely oriented around the project of giving a normative theory for actions under moral uncertainty. To illustrate, suppose you know that you have some minor moral reason in favour of ordering veal, and think it slightly more likely than not that ordering veal would be morally blameless because the interests of veal calves do not matter morally. But you also think it only slightly less likely than not that the interests of veal calves matter a great deal, such that ordering veal would be morally on a par with committing murder. Most participants in the literature think what you ought to do in this situation depends at least in part on your divided moral beliefs. One family of theories prescribes morally hedging, that is, not ordering veal because doing so would be morally reckless. An example of such a hedging theory would be what we may call 'Expected Moral Value Theory' or 'EMVT'. According to EMVT, an agent should maximize their expected moral value, which in this case would mean not ordering the veal.[1] Another family of theories recommends ordering the veal, because moral hedging requires the

[1] Advocates of EMVT include Jacob Ross, 'Rejecting Ethical Deflationism', *Ethics* 116 (2006), pp. 742–68, and Andrew Sepielli, 'What to Do When You Don't Know What to Do'. *Oxford Studies in Metaethics* 4 (2009), pp. 5–28. Advocates of hedging more broadly construed include Graham Oddie, 'Moral Uncertainty and Human Embryo Experimentation', *Medicine and Moral Reasoning*, ed. K. Fulford (Cambridge,

Utilitas Vol. 30, No. 4, December 2018
doi:10.1017/S0953820818000080

agent to make inter-theoretic value comparisons, a task that they cannot non-arbitrarily complete.[2] An example of such a non-hedging theory would be My Favourite Theory (MFT). MFT prescribes the action with the highest value according to the theory the agent has the highest credence in. In this case that means ordering veal.[3]

But a third kind of theory rejects the entire project of offering a normative theory for moral uncertainty. This approach, called ‘Normative Externalism’ by Brian Weatherson, and ‘Actualism’ by Elizabeth Harman, rejects internalist factors like moral beliefs and accessible moral evidence as irrelevant to an agent’s moral reasons – an agent is morally required to do what morality actually requires of them, their moral uncertainty notwithstanding.[4] Attempts to craft theories in response to moral uncertainty are fundamentally misguided on this view. I will follow Weatherson in calling this view ‘Normative Externalism’ or simply ‘Externalism’, and the various theories that consider moral uncertainty to be morally salient as ‘Normative Internalism’ or simply ‘Internalism’.

Externalists have levelled a couple of arguments against Internalist theories. One, advanced by Weatherson, accuses Internalism of requiring agents to be moral fetishists, because these theories all require an agent to care about *doing the right thing* read in a *de dicto* sense.[5] Weatherson argues that this kind of moral motivation is fetishistic and thus an embarrassment for Internalism. In this article, I focus on a second argument against Internalism, raised by Harman, which

1994), pp. 144–61 and Alexander Guerrero, ‘Don’t Know, Don’t Kill: Moral Ignorance, Culpability, and Caution’, *Philosophical Studies* 136 (2007), pp. 59–97.

I use the expression ‘moral value’ in a way that is meant to be neutral with respect to all first-order moral theories and their axiological commitments or lack thereof. To the extent that a moral theory makes value comparisons, such that one action is better or worse or equal to another, that theory assigns *moral values*, on my use of the term. Some authors prefer the expression ‘moral choice-worthiness’ instead of ‘moral value’ in an attempt to avoid any confusion on these grounds. See for example W. MacAskill, ‘Moral Uncertainty as a Voting Problem’, *Mind* 125 (2016), pp. 967–1004. For purely aesthetic reasons, I opt for ‘moral value’ instead of the more cumbersome ‘moral choice-worthiness’.

[2] To hedge one must consider the potential disvalue of wrongly choosing the action you believe is probably right. But this consideration requires making inter-theoretic value comparisons – you must have a sense that it would be much *worse* to φ if one theory (the one you think less likely) is true than to not φ if another theory (the one you think more likely) is true.

[3] MFT is advocated by Edward Gracely, ‘On the Non-Comparability of Judgments Made by Different Ethical Theories’, *Metaphilosophy* 27 (1996), pp. 327–32, and Johan Gustafsson and Olle Torpeman, ‘In Defence of My Favourite Theory’, *Pacific Philosophy Quarterly* (2014), pp. 159–74. Other non-hedgers include MacAskill, ‘Voting Problem’.

[4] Brian Weatherson, ‘Running Risks Morally’, *Philosophical Studies* 176 (2014), pp. 141–63; Elizabeth Harman, ‘The Irrelevance of Moral Uncertainty’, *Oxford Studies in Metaethics* 10 (2015), pp. 53–79.

[5] Weatherson, ‘Running Risks’.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

I will refer to as the 'exculpation problem'.[6] According to Harman, Internalism entails that moral ignorance exculpates the agent who has acted (objectively) wrongly out of her false moral beliefs, because such wrongful actions would be prescribed by Internalism. Harman then argues that moral ignorance does not exculpate, citing as evidence cases in which our intuitions are that such agents are indeed culpable.

I argue that Harman's objection fails, because although there is a version of Internalism against which the exculpation problem is sound, it is not a version that any Internalist defends, at least not in print. That version of Internalism is one that issues *moral* prescriptions based at least in part on an agent's *subjective* probability function over first-order moral theories. If Internalism is instead construed either as a theory that issues rational prescriptions or as a theory that considers only an agent's *epistemic*, or evidence-based, probability to be salient, then the exculpation problem misses its mark.[7] Though Harman's objection misses its mark, it raises important issues about how Internalism should be understood. I argue for a rational reading of Internalism's prescriptions on the grounds that it avoids what I will call the 'deontic conflict problem', a problem that seems decisive against the moral reading of Internalism's prescriptions.

II. THE EXCULPATION PROBLEM

Harman argues that Internalism entails that moral ignorance is morally exculpatory for the agent. While different Internalist theories will disagree about which elements of an agent's belief structure are salient, they all entail that an agent should act on her false moral beliefs in cases in which an agent is certain that some theory, T1, is true and that T1 prescribes φ-ing, but φ-ing is objectively wrong.[8]

[6] Harman, 'Irrelevance'.

[7] Authors have in fact independently advanced each of these responses. Andrew Sepielli, 'How Moral Uncertaintism can be both True and Interesting', *Oxford Studies in Normative Ethics* 7 (2017), pp. 98–116, responds to the exculpation problem by appealing to epistemic probabilities, while Krister Bykvist, 'Evaluative Uncertainty, Environmental Ethics, and Consequentialism', *Consequentialism and Environmental Ethics*, ed. A. Hiller, R. Ilea and L. Kahn (New York, 2014), pp. 122–35, appeals to weakening the normative domain of Internalism's prescriptions from the moral domain to the instrumentally rational domain (Bykvist's paper was published a year before Harman's and thus does not address *her* argument, but one similar enough for all intents and purposes). Though these authors' arguments share certain core similarities to my response to Harman, I offer some important supplements and criticisms of their work. My article also takes a wider view of the problem, encompassing both of these responses, and ultimately offering novel substantive arguments for the rational reading of Internalism.

[8] It may be objected that Harman is using a case of moral certainty to impugn theories for moral *uncertainty* – perhaps Internalist theories are meant to be silent on such cases. We can easily modify the case to navigate this objection. Suppose instead that the

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

If T1 prescribes φ-ing, then every Internalist theory will likewise prescribe φ-ing, though they will do so for different reasons – EMVT will prescribe φ-ing because it maximizes expected moral value, for example, while MFT will prescribe φ-ing because it is the action with the highest moral value according to the theory the agent has the highest credence in. If an agent cannot be culpable for doing what they ought to do, and if Internalism says an agent ought to φ on the basis of their moral beliefs, then Internalism entails that moral ignorance exculpates.

Harman then argues that moral ignorance does not exculpate.[9] She offers two cases in which an agent is certain that φ-ing is morally obligatory, but in which it is in fact seriously wrong. Consider:

> *Max*
> Max works for a Mafia 'family' and believes he has a moral obligation of loyalty to the family that requires him to kill innocents when it is necessary to protect the financial interests of the family. This is his genuine moral conviction, of which he is deeply convinced. If Max failed to 'take care of his own' he would think of himself as disloyal and he would be ashamed.

> *Gail*
> Gail is a gang member who believes that she has a moral obligation to kill a member of a neighbouring gang as revenge after a member of her own gang is killed, although her victim was not responsible for the killing. This is her genuine moral conviction, of which she is deeply convinced. If Gail failed to 'take care of her own' she would think of herself as disloyal and she would be ashamed.[10]

According to Harman, Max and Gail are both 'paradigm cases of agents blameworthy for their wrongful actions'.[11] Harman argues that even if we fill in Max's and Gail's stories with more detail to the effect that they are justified in holding their false moral beliefs, they would still strike us as morally culpable.

agent is morally uncertain, but her uncertainty is structured such that φ-ing strongly dominates all other actions. That is, according to every theory she is considering, φ-ing is better than all other actions. Every Internalist theory will prescribe φ-ing under these conditions. Now suppose that φ-ing is objectively wrong. In this case, all Internalist theories will direct a morally uncertain agent to act on her false moral beliefs.

[9] Harman argues for this claim first in 'Does Moral Ignorance Exculpate?', *Ratio* 24 (2011), pp. 443–68, later applying it against Internalism in 'Irrelevance'.

[10] Harman, 'Irrelevance', p. 65.

[11] Harman, 'Irrelevance', p. 65.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

III. WHICH VERSION OF INTERNALISM – RATIONAL OR MORAL PRESCRIPTIONS?

Does Internalism entail that Max and Gail are morally exculpated? I argue that the answer depends on which version of Internalism one has in mind. Besides the distinction between hedging and non-hedging theories, Internalist theories can be individuated along two additional dimensions. First, Internalism can issue either moral prescriptions or instrumentally rational prescriptions. In other words, we can ask whether Max and Gail are morally required to act on their beliefs, or whether they are (merely) rationally required to, according to Internalism. The second dimension along which Internalist theories may individuate themselves is whether it is an agent's subjective probabilities or their epistemic probabilities that are salient to the question of what they ought to do. An agent's subjective probabilities are just the distributions of split belief that an agent actually has over various mutually exclusive and jointly exhaustive sets of possibilities, irrespective of their evidence. Epistemic probabilities, on the other hand, are roughly the distributions that the agent epistemically ought to have, given their evidence. In other words, is it just Max and Gail's actual graded beliefs about which moral theory is true that matter for Internalism, or is it the graded beliefs they would have, were they forming their beliefs in epistemic compliance with their evidence? In this and the following section I will argue that the exculpation problem is only a problem for a version of Internalism that issues *moral* prescriptions in virtue of an agent's *subjective* probabilities. But in fact no proponent of Internalism defends this view in print (or in person as far as I know).

Let us begin by considering the distinction between rational and moral obligations. A key premise of the exculpation problem is that an agent cannot be culpable for doing as they ought. Harman has the following sort of principle in mind:

1. If one ought to φ, then one is not culpable for φ-ing.[12]

But this principle is not precise enough because it leaves open which normative domain 'ought' and 'culpable' belong to. Here is a way of filling this in that is obviously false:

2. If one prudentially ought to φ, then one is not morally culpable for φ-ing.

[12] Harman asserts roughly the contrapositive of 1: 'An agent is blameworthy for her behavior only if she acted as she subjectively should not have acted' ('Irrelevance', p. 56). Harman uses 'blameworthy' and 'culpable' interchangeably, and her wording makes explicit what I leave implicit – that the 'ought' is subjective. I take 1 to be equivalent to Harman's formulation.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

According to 2, a murderer who acted prudently (say the murder was in her interests and she was careful not to get caught) is thus not morally culpable for murdering. But this is absurd.[13] Here is a version of the principle that adjusts for this concern and seems very plausible:

3. If one ought to φ according to some norm, N, then one is not culpable for φ-ing with respect to N.

This seems to be the sort of principle at work in Harman's argument that Internalism entails that moral ignorance morally exculpates. But in that case, if Internalism issues instrumentally rational prescriptions, then the exculpation problem is unsound as one can do as one rationally ought without morally exculpating oneself.

Do Internalists understand their theory as issuing moral or rational prescriptions? The moral reading of Internalist prescriptions is a natural one, as Internalists often seem to describe compliance and non-compliance with their theories in moral terms. For example, Graham Oddie, a proponent of hedging, argues that lethally experimenting on human embryos is 'morally justified' only if the goods obtainable by such experiments are likely to be considerable – otherwise even a very small probability that the embryos have the same moral status as us would rule out such experimentation.[14] Similarly, Johan Gustafsson and Olle Torpman, non-hedging proponents, frame their question as asking which actions would amount to a 'morally conscientious choice' for a morally uncertain agent.[15] But just as often these failures are *explained* in the language of instrumental rationality. Both Oddie and Gustafsson and Torpman proceed to explain and defend their view in highly decision-theoretic language suggestive of the normative domain of instrumental rationality.

Still other authors are explicit that they mean their Internalist theories to be issuing rational prescriptions. Jacob Ross, for example, writes: 'So long as the various ethical theories in which we have credence can be given an appropriate quantitative representation, it will be possible to employ decision theory in determining what choices would be most *rational* under ethical uncertainty.'[16] Likewise, Ted Lockhart is quite clear throughout his book on the subject

[13] I do not mean to suggest that Harman asserts a general principle like 1, while failing to appreciate the danger of making it precise in the manner of 2. In fact, Harman specifically addresses the worry that Internalism's prescriptions might not be moral, which I will address shortly. I only introduce 2 to underscore the fact that 'oughts' and culpability are related in such a way that for an entailment like 1 to go through, they must agree with respect to their normative domain.

[14] Oddie, 'Embryo'.

[15] Gustafsson and Torpman, 'Defence'.

[16] Ross, 'Deflationism', p. 755 (my italics).

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

that his normative theory for moral uncertainty places rational, not moral, constraints on morally uncertain agents.[17] Ralph Wedgwood, in a paper offering a general principle describing the irrationality of *akrasia*, ends up in a position which is basically EMVT. He argues that '*rationality* requires one to have an intention that . . . maximizes expected choiceworthiness' and argues that this principle governs decision-making under moral uncertainty as well as non-moral uncertainty.[18]

Krister Bykvist similarly argues that uncertain agents have *rational* reasons to act in light of their uncertainty, but not moral reasons.[19] Bykvist's argument anticipates Harman's paper, arguing for rational instead of moral reasons on the grounds that doing so would avoid counter-intuitive implications similar to the ones Harman raises. More extreme than Max and Gail, Bykvist discusses a case of an extreme egoist who tortures children for fun because he believes this is what egoism requires. He is, Bykvist stipulates, justifiably certain that egoism is true, and that it implies that he is morally justified in torturing children. Bykvist thinks the torturer is rationally, but not morally exculpated in light of his justified moral beliefs. In this case, more precisely, it is his axiological beliefs that rationally justify his actions – Bykvist's argument takes place in the narrower context of axiological uncertainty for consequentialists, not the broader moral uncertainty debate. But his argument can easily be generalized.

So, a number of Internalists clearly endorse a rational reading of Internalism's prescriptions. But perhaps they are mistaken. Perhaps Harman's exculpation problem can be understood as objection to the best understanding of Internalism, and these authors' positions can be rebutted on the independent grounds that a rational understanding of Internalisms' prescriptions is wrongheaded. Harman in fact offers an argument for the moral understanding of Internalisms' prescriptions. Harman asks us to consider someone named 'Bill' who simply does not have any moral goals, but still has some moral beliefs.[20] Harman supposes that Internalists will still want to say that Bill should act on those beliefs in an appropriate way. For example, advocates of EMVT will say that Bill ought to avoid being morally reckless. But if EMVT's prescriptions belong to the domain of instrumental rationality, this normative constraint on Bill is unjustified. Bill would not be

[17] Tedd Lockhart, *Moral Uncertainty and its Consequences* (Oxford, 2000).

[18] Ralph Wedgwood, 'Akrasia and Uncertainty', *Organon* 20 (2013), pp. 484–506, at 494–5 (my italics).

[19] Bykvist, 'Evaluative Uncertainty'.

[20] Harman, 'Irrelevance' p. 55.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

instrumentally irrational if he acted morally recklessly because he has no moral goals.

Internalists who endorse a rational reading of the theory's prescriptions may very well want to say that Bill ought to, say, avoid moral recklessness, but I argue that this judgement can be made sensible on the rationality reading. A plausible *corollary* to the rational reading of Internalism, though not necessarily a part of the theory itself, is that one ought to have some morally appropriate moral goals. There is no need here to be dogmatic about exactly which goals are morally appropriate. Some are clearly inappropriate, like a goal of maximizing suffering. Other goals will presuppose which Internalist theory an agent should act under. For example, if my goal is to act in compliance with the theory that is most probably true, irrespective of moral peril, then I rationally ought to act in accordance with MFT, not EMVT. Adequately unpacking the issue of which goals are appropriate would require more extensive treatment than I can provide in this article, but all I need to disarm the present objection is the fairly uncontroversial claim that morally virtuous agents operate with certain moral goals in mind – they aim at the good. An agent, like Bill, who lacks any such goal, is in some sense morally deficient. He ought to repair this deficiency. To this independently plausible claim, the Internalist adds that, with this deficiency repaired, Bill ought to act rationally with respect to his moral goals. So, saying that Bill ought to avoid moral recklessness could just be understood as shorthand for, 'Bill ought to have such and such moral goals, and if (and only if) he has these goals, he ought to act rationally with respect to them'.

Consider an analogous case involving welfare. Some welfare theorists believe the right theory of welfare to be a hybrid theory, conjoining desire satisfaction theory and objective list theory. On this view, something is good for someone if and only if they desire it *and* it is objectively good. Suppose one of the objectively good things is friendship. Now imagine someone, call him 'Brad', who does not desire friendship and has no friends. We can imagine Brad's concerned family, all of whom are committed hybrid theorists about welfare, saying to themselves 'it would be good for Brad to have friends'. If we confronted them by pointing out that a necessary condition for it to be good for Brad that he have friends was not satisfied, namely his having a desire for friends, it seems they would be justified in responding, 'yes, yes – we only mean that it would be good for him to have friends *and* desire to have them'.

So it is for the rational reading of Internalism and Bill. Proponents of this theory may sensibly say that Bill ought to hedge, and by that mean more precisely that Bill ought to hedge if and only if he has certain moral goals *and* that he ought to have those goals. The second

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

conjunct of this claim about Bill must be justified independently, but I take it that the more minimal claim that an agent morally ought to have *some* moral goals and that some other moral goals are morally out of bounds is quite plausible on its face. Justifying a more precise and theoretically partisan goal, such as one amenable to EMVT or MFT, will require additional arguments, which I will not attempt to provide here. The upshot is that the rational reading seems open for Internalists to take, and some Internalists, such as Ross and Lockhart, do in fact take it. But Harman's exculpation objection misses its mark against this version of Internalism.

IV. WHICH VERSION OF INTERNALISM – SUBJECTIVE OR EPISTEMIC PROBABILITIES?

Now let us consider probability. If we assume, contrary to some proponents of the theory, that Internalism issues moral prescriptions, then it seems that if Internalism prescribes φ-ing, φ-ing is morally blameless, contrary to our intuitions about Max and Gail. But I will argue that this is not quite right. More specifically, I will argue that our intuitions about Max and Gail are tainted by their implicit lack of epistemic justification for their beliefs. The upshot is that if Internalism is characterized in such a way that only epistemic probabilities are salient to the decision analysis, then Internalism's prescriptions will not clash with our culpability-finding intuitions, even if those prescriptions are moral prescriptions.

This is more or less the move Andrew Sepielli makes in response to Harman. Sepielli argues that his version of Internalism 'affords no right-making role to the agent's credences' (by which he means subjective probabilities), but only to the agent's epistemic probabilities. His version of Internalism does not entail the embarrassing consequence that agents like Max and Gail are exculpated because the moral theories they subscribe to are not epistemically justified. However, Sepielli fails to address Harman's claim that agents like Max and Gail are culpable *even if their beliefs are epistemically justified*, though he acknowledges that 'Harman may well want to reject norms that are relative to the epistemic probabilities of moral claims, too'.[21] Perhaps Sepielli thinks, as I do, that Harman's claim against the exculpation of even epistemically justified agents is under-argued, and so thinks this is a reasonable place to leave the dialectic. I will go further, arguing that the cases of Max and Gail give us no moral data with respect to the culpability of epistemically justified agents because

[21] Sepielli, 'Uncertaintism', p. 103.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

our culpability-finding intuitions in these cases are entirely due to their implicit lack of epistemic justification.

Even if we stipulate, as Harman does, that Max and Gail are epistemically justified, it will be difficult as readers to blind ourselves from what will probably seem to us to be a glaring lack of justification. After all, what could justify someone believing that a mob hit or a tit-for-tat gang killing is morally obligatory? To blind ourselves in such a way as to make an assessment that is not tainted by considerations of epistemic justification, we need alternative cases in which the background details that implicitly preclude epistemic justification are removed. One such example would be a case involving a more ethically contentious choice, in which it is plausible that an agent's moral beliefs are justified, though wrong. Another would be a formalized case devoid of any potentially intuition-skewing details about the content of the agent's moral beliefs. I offer one of each for good measure.[22]

First, consider the case of Glinda. Glinda is pregnant and is considering having an abortion. She diligently researches the matter, reading the relevant ethics literature, conversing with experts with differing opinions, and reflecting carefully on her own beliefs in an attempt to root out any biased, incoherent or unjustified beliefs. By the end of this process Glinda has become fully convinced that there is nothing wrong with having an abortion. Let us stipulate that she is mistaken about this, and that having an abortion in her circumstances is in fact seriously wrong, as wrong as killing an innocent person.[23] Is Glinda morally culpable for acting out of her false moral beliefs by having an abortion? I am strongly inclined to say that she is not, and I hope the reader will agree.

[22] A third strategy might be to construct a plausible and more robust epistemic backstory for both Max and Gail such that their moral ignorance does indeed seem justified. Suppose they were raised in a family and culture in which mob hits and revenge killings were widely believed to be justified, suppose all of the most articulate and authoritative sources of moral knowledge available to them also held these beliefs, and so on. I doubt that we could completely expunge our strong moral aversion to mob hits and revenge killings this way, but for what it's worth, I find my culpability-finding intuitions weakening considerably the more robust this backstory becomes. This, of course, tells in favour of my claim that our intuitions are tracking a lack of epistemic justification. Each of these strategies reinforces the others and underscores the flaws in Harman's original intuition pumps.

[23] I assume that abortion is an issue about which reasonable and well-informed people may disagree about the moral facts. It does not really matter for my purposes what the moral facts are about abortion. We could switch the details and make Glinda confident that abortion is wrong, when in fact it is not (adding the supposition that Glinda has some very strong moral reason to have the abortion – perhaps because attempting delivery would carry with it a high risk of maternal death, leaving Glinda's other children without a mother). I take it we would still find Glinda not culpable for whatever actions she takes so long as we also believe her to be epistemically justified.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

Now consider Matt. Matt is considering φ-ing. He diligently researches the matter, reading the relevant ethical literature, conversing with experts with differing opinions, and reflecting carefully on his own beliefs in an attempt to root out any biased, incoherent or unjustified beliefs. By the end of this process Matt has become fully convinced that there is nothing wrong with φ-ing. Suppose φ-ing is actually seriously wrong, as wrong as killing an innocent person. Is Matt culpable for φ-ing? Again, I am strongly inclined to say that he is not.

The cases of Matt and Glinda are structurally identical to the cases of Max and Gail, involving similar moral peril (wrongfully killing innocent people, or a morally equivalent action). But unlike revenge killings, the moral status of having an abortion seems like the sort of thing people could be justifiably mistaken about. And unlike a mob hit, the moral status of φ-ing, because it is left vague, seems like the sort of thing that, *as far as we know*, someone could be justifiably mistaken about. All of this suggests that the intuitions we have in the Max and Gail cases are tainted by their implicit lack of epistemic justification.

This shows that Harman's cases fail to establish what she wants them to, namely that moral ignorance, *even when it is justified*, does not morally exculpate.[24] If that is the case, then the epistemic probability reading of Internalism is still in business, even if it is paired with a moral reading of Internalism's prescriptions. The intuitions generated by the cases Harman offers simply do not count as data against the claim that justified moral ignorance exculpates. And it would be difficult if not impossible to offer other cases that would satisfactorily avoid the worry about tainted intuitions. Any case that features an Internalism-compliant agent who is putatively culpable will fall into one of two traps. Either our culpability-finding intuitions will be very strong, but will be paired with implicit incredulity about the agent's epistemic justification in believing that she is morally permitted to act as she is. Or the agent will be believably justified in her moral ignorance, as in the cases of Glinda and Matt, but we will fail to have culpability-finding intuitions. In the first case, the epistemic probability Internalist can simply reject that their theory has the

[24] The failure of these cases extends beyond Harman's 2015 paper against Internalism, to her earlier paper, 'Does Moral Ignorance Morally Exculpate?' *Ratio* 24, (2011), pp. 443–68. In this paper, Harman is responding to sceptical worries raised by Gideon Rosen ('Skepticism about Moral Responsibility', *Philosophical Perspectives* 18 (2004), pp. 295–313) about the extent to which anyone is really morally culpable for anything. Harman argues that even a narrower thesis of Rosen's, roughly that *justified* moral ignorance exculpates, tells against our intuitions about culpability. Harman cites cases similar to Max and Gail along with several others. The cases of Matt and Glinda should undermine our intuitions in these cases as well.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

perverse implication, as Sepielli does. In the second case, they may simply shrug their shoulders and embrace the no-longer-embarrassing implications. Bykvist's child-torturing egoist case is punctured on the first horn of this dilemma. Any version of egoism that entails the torturer's exculpation is a version that no reasonable person could believe in. If that is what egoism entails, then egoism is not only a false theory, but an epistemically unjustified one.

V. THE DEONTIC CONFLICT PROBLEM

I have argued that there is only one version of Internalism susceptible to Harman's exculpation problem – one that pairs subjective probabilities with moral prescriptions. I know of no one who explicitly defends such a version of Internalism, but suffice it to say the exculpation problem renders such a view untenable. While it is not successful as a refutation of Internalism, the exculpation problem is helpful as a prompt for Internalists to get clearer on precisely what their theory is. They must choose between two options, either (1) a theory that issues instrumentally rational prescriptions in light of an agent's subjective or epistemic probabilities ranging over first-order moral theories, or (2) a theory that issues moral prescriptions in light of an agent's epistemic probabilities ranging over first-order moral theories. In this section I argue against the second, moral, reading on account of what I will refer to as the 'deontic conflict problem'.

The deontic conflict problem arises from the fact that Internalism will routinely issue prescriptions that contradict one, and sometimes contradict *all*, of the first-order moral theories among which the agent's beliefs are split. If Internalism's prescriptions are subjective moral prescriptions, this raises some uncomfortable questions for the theory. As a striking example of deontic conflict, consider an agent, call her 'Angela', whose moral beliefs are divided between two or more theories that are satisficing in nature. Angela is considering whether to abort an unwanted pregnancy. She is completely confident that foetuses have a right to life, and her beliefs are split between two rights-based moral theories, according to either of which it would be morally heroic, but not morally required, to preserve the life of an innocent being with a right to life at great personal cost (such as the cost of carrying a pregnancy to term and delivering a baby).[25]

According to all of Angela's moral beliefs, having an abortion in her circumstances is morally permissible, but according to either

[25] The inspiration for this case is obviously Judith Thomson's famous paper 'A Defense of Abortion', *Philosophy and Public Affairs* 1 (1971), pp. 47–66.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

MFT or EMVT, it is prohibited. Continuing the pregnancy is the action with the highest moral value according to either theory Angela has credence in. Both of those first-order theories consider the action to be supererogatory, but neither MFT nor EMVT countenance supererogatory actions as a normative category. You must perform the optimal action according to either of these versions of Internalism. And the optimal action, heroic by Angela's lights, is to complete the pregnancy and deliver the baby.

This case is provocative for either the moral or rational readings of Internalism, but I will argue that it only constitutes an objection to the moral reading. If Internalism is understood as issuing moral prescriptions, then this prohibition on abortion in Angela's case is a subjective moral prohibition – one that *runs contrary to all of Angela's first-order moral beliefs*. This is more than a little strange. For Internalism to contradict the unanimous moral opinion of the first-order theories it is meant to adjudicate among, it must be generating an independent moral maxim – maximize expected moral value, or do what is best according the theory you are most confident in. As the case of Angela shows, these moral maxims do not just fall out of the agent's first-order beliefs. They require independent justification. But there is no way to argue for the independent justification of these moral maxims that does not beg the question against any number of first-order moral theories. In our example, justifying EMVT or MFT as moral maxims requires Internalism to take a position against the category of supererogatory actions, and thus against most deontological theories. This general problem holds no matter which second-order decision rule we attempt to justify for Internalism. A satisficing rule might dodge the issue in the case of Angela, but not in a case with an agent whose credence is divided over various maximizing first-order moral theories. If the rule constitutes a moral requirement, then it presumes an answer to matters about which Internalism is supposed to remain neutral, an answer that is supposed to be provided by the first-order moral theories.

This is a problem if Internalism is a theory that issues moral prescriptions, but not if it issues rational prescriptions. The rational version of Internalism does not presume to answer moral questions. It does make substantive claims about what is rationally required, but not about what is morally required. By choosing an action that she knows has less value according to either theory she is uncertain about, it is perfectly plausible to say that Angela would be in some sense acting irrationally if she has the abortion, even if she would not be acting wrongly in the subjective moral sense. These substantive claims about Angela's rational requirements, whatever they end up being, can be justified independently without encroaching on any claims

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

made by first-order moral theories. One can act morally without acting rationally, in this case because the requirements of rationality are more stringent than those of subjective morality.

So, if Internalism is understood as issuing moral prescriptions, then it makes substantive, independent moral claims that require independent justification. To my knowledge, no one in the literature attempts to do this, and for good reason – it would betray the shared sense that Internalism is meant to be a neutral theory for moral uncertainty, not a partisan player in normative ethics. But without such independent justification, these moral maxims arbitrarily bias Internalism's decision analysis in favour of first-order theories that have similar decision-guiding principles. Because the rational reading of Internalism's prescriptions avoids this implication, it should be favoured.

VI. OBJECTIONS

I will briefly consider two objections to what I have argued in this article. Both objections are inspired by remarks Harman makes in response to a proposal not entirely unlike the proposal I have advocated, that Internalism be understood as issuing rational prescriptions. And both objections involve the charge that this understanding of Internalism would be in some sense uninteresting. According to the first objection, the rational reading of Internalism makes the theory *philosophically* uninteresting. Because the rational understanding might be thought to have limited the ambition of the Internalist project, such that there is a sense in which Max and Gail do as they ought (a rational sense), and also a sense in which they do not (a moral sense), it ends up making true claims about Max and Gail (that they do as they ought), but only in an obvious and philosophically uninteresting way. The second objection is that by no longer issuing moral prescriptions Internalism is not only theoretically uninteresting, but also morally inert, and thus *practically* uninteresting. This would be troubling, for example, if it left hedging theories without the resources to make sense of the charge of moral recklessness, a charge widely understood by proponents of hedging theories to be a kind of *moral* wrongdoing.

Harman considers several possible responses to the exculpation problem, one of which sounds similar in many respects to understanding Internalism's prescriptions as rational prescriptions.[26] The response Harman considers also involves the charge of equivocating on the sense of 'ought' and 'culpable' in the conditional 'If one ought

[26] Harman, 'Irrelevance', pp. 72–3.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

to φ, then one is not culpable for φ-ing'. But the kind of equivocation she considers is between a sense of 'ought' that is relative to all of an agent's beliefs and a sense of 'ought' that is sensitive only to an agent's moral beliefs. Harman thinks Externalists should grant that there is a sense in which Max should carry out the mob hit. Relative to *only his moral beliefs*, it is true that he should. However, if this is what Internalism is claiming, it is an uninteresting claim, similar to saying of someone, Nora, who has been told some falsehood, P, that she should be able to easily recognize as false, that Nora should believe P, *if her beliefs should be formed on the basis of their interlocutor's testimony alone*. Once this conditional is added, then the claim is obviously true, but also uninteresting. If Internalism is merely claiming that Max should carry out the mob hit, *if his moral beliefs alone are pertinent*, then Internalism is likewise making a true but uninteresting claim about Max.

Rather than limiting the set of beliefs relative to which Internalism makes prescriptions, part of what I have argued could be construed as limiting the normative domain in which Internalism's prescriptions are made. Maybe an objection similar to Harman's applies to this limiting move as well. Perhaps it is obviously true that Max *rationally* ought to carry out the mob hit given his beliefs, and thus uninteresting that Max ought to carry out the mob hit, if by 'ought' we mean 'rationally ought'.

Limiting the normative domain of Internalism does not make it theoretically uninteresting in the same way that limiting the set of beliefs relative to which it makes prescriptions might make it theoretically uninteresting. There is simply nothing tautologous or obvious about the claim that an agent rationally ought to, say, maximize expected moral value. If there were, then the ongoing debate between hedging and non-hedging theorists would make no sense if the competing families of theories are both issuing rational prescriptions. But the debate between the two kinds of Internalism is perfectly sensible on either the moral or rational reading of the theories' prescriptions. On the rational understanding of Internalism, the debate may boil down to a dispute about which goals and corresponding decision rules morally uncertain agents should employ, but this is still an important and interesting debate, on which much hangs. For one thing, the goals and rules favoured by non-hedging theorists allow them to skirt the problem of inter-theoretic value comparison, thought to be a significant obstacle for hedging theories like EMVT.[27]

[27] See e.g. Gustafsson and Torpman, 'Defence' and MacAskill, 'Voting Problem' on how non-hedging theories have an advantage on this issue.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

Hedging theories, on the other hand, seem to do a better job matching our intuitions about how agents ought to act. Non-hedging theories, because they avoid making inter-theoretic value comparisons, often make surprising and implausible act prescriptions. This all strikes me and, I assume, the Internalists who are engaged in this debate, as very theoretically interesting.[28]

The second way in which the rational reading of Internalism might be uninteresting is by being practically uninteresting. If Internalism issues rational, not moral prescriptions, then perhaps it is irrelevant to our practical deliberations. Just as Nora should not care about what she ought to believe in the limited sense considered, so too morally uncertain agents should not care about what Internalism prescribes. The basic worry is that the norm of instrumental rationality simply does not carry the kind of normative force or urgency that the norm of morality does. This is especially worrying in light of the fact that most advocates of Internalism clearly want the theory to do some moral work. As I described earlier, Internalists often characterize the issue of moral uncertainty as what an agent is morally required to do, even if their explanations of those moral requirements invoke the language of instrumental rationality.

Although compliance with the rational understanding of Internalism does not entail moral exculpation, this does not mean the theory has no moral implications at all. On the contrary, there is some independent plausibility to the notion that acting rationally with respect to one's moral beliefs and moral goals is *necessary* for doing one's subjective moral best. There are many ways in which an agent might fail to do their subjective moral best. For example, they could fail to have justified moral beliefs, like Max and Gail, or Bykvist's child-torturing egoist. Or, they could hold grossly inappropriate moral goals, like maximizing suffering, or fail to have any moral goals at all, like Bill. But they could also fail to act rationally with respect to their moral beliefs and moral goals, and it is this kind of failure for which Internalism is offering a normative theory.

I take it this analysis coheres with our intuitions about Angela. On this analysis, we may say that Angela does nothing subjectively morally *wrong* if she has the abortion – she has, after all, acted in moral compliance with her justified first-order moral beliefs – but she does fail to act *rationally* with respect to her moral beliefs, according

[28] I again want to stress that Harman herself is not objecting to the rational reading of Internalism on these grounds. The target of her objection is closely enough related to the rational reading that it is worth considering, but I do not want to falsely attribute a weak objection here to Harman. Her worry about interestingness does strike me as a sound objection against the target she has in mind.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

to either MFT or EMVT. Because of this, she fails to do her subjective moral *best*. This should hardly be controversial. Both of the moral theories Angela's credences were split between agree that she could do better – she could have acted heroically. She is not morally culpable and perhaps she is not even morally criticizable, but she still fails to perform the morally optimal action. And this failure is explained by her non-compliance with Internalism (assuming some optimizing version of Internalism, like MFT or EMVT, is true).

Under certain circumstances, failing to act rationally in light of one's moral beliefs and goals can amount to the *moral* failure of moral recklessness. That is, under certain conditions, an agent's failure to act rationally can amount to full-fledged moral wrongdoing and corresponding moral culpability. In our opening case of ordering veal, for example, this seems to be a plausible analysis, at least by the hedging theorist's lights. If you order veal when you believe that ordering a salad instead would incur almost no normative cost, moral or otherwise, and despite the fact that you think there is a significant probability that ordering veal is morally heinous, akin to murder, then you have done something subjectively morally wrong. And the source of your moral wrongdoing was a failure to comply with certain norms of instrumental rationality relative to some moral goals.

It is not terribly important here to spell out exactly which conditions must be satisfied for mere rational failures to double as moral failures. It is enough to establish that sometimes this transformation does happen, and paradigm cases of moral recklessness are such occasions. That the transformation from rational wrongdoing to moral wrongdoing happens under conditions of recklessness is attested to by everyone in the literature, at least when the underlying uncertainty is non-moral in nature. No one in the literature denies that it is morally wrong to feed a cake to a guest when one thinks there is a non-negligible chance that you added poison instead of vanilla.[29] And all agree that this act would be wrong even if the cake is in fact poison-free. The wrongness is accounted to the recklessness of the action. But recklessness is just a kind of instrumental irrationality occurring under circumstances involving highly asymmetric potential peril. If the transformation from irrational action to immoral action uncontroversially occurs for poisoned cake cases, then there is no non-question-begging reason why it could not also occur in veal cases as well, or for cases of moral uncertainty more generally.

[29] I borrow this case from Weatherson, 'Running Risks'.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press

VII. CONCLUSION

I have argued that Harman's exculpation problem misses its mark. It is sound only against a version of Internalism that combines moral prescriptions with subjective probabilities, a combination that no one endorses in print. I have also argued that the normative domain in which Internalism issues prescriptions is best understood as the domain of instrumental rationality, because the moral version of Internalism fails to navigate the deontic conflict problem successfully.

Some daunting problems remain for Internalism. For hedging theories, there is the problem of inter-theoretic value comparison.[30] For Internalism more generally, there is the charge that the theory requires agents to be moral fetishists.[31] And for the rational reading of Internalism, much more needs to be said about how to think about moral goals – for example, whether there is a single kind of moral goal that is appropriate, or whether Internalists should be ecumenical on this issue. I think Internalists should take heart. The literature is still young and by my lights no appealing alternatives have presented themselves.[32]

Jay.Geyer@Colorado.edu

[30] See e.g. Ross, 'Deflationism', and Brian Hedden, 'Does MITE Make Right?', *Oxford Studies in Metaethics* 11 (2016), pp. 102–28.

[31] Weatherson, 'Running Risks'.

[32] This article has benefited from the thoughtful comments of many readers and audience members, including those at the University of Colorado Boulder and at the 2017 Eastern APA. I especially want to thank Graham Oddie, Jonathan Spelman, Macy Salzberger, two anonymous reviewers for this journal, and Krister Bykvist for their very helpful written comments.

https://doi.org/10.1017/S0953820818000080 Published online by Cambridge University Press