Spring 1-1-2018

Reasons and Value: Decision-Making Under Moral Uncertainty

Jay R. Geyer
University of Colorado at Boulder, jaygeyer3@yahoo.com

REASONS AND VALUE: DECISION-MAKING UNDER MORAL UNCERTAINTY

By

JAY R. GEYER

BA, University of Illinois, 2010

MA, University of Colorado, 2013

A thesis submitted to the
faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Philosophy
2018

This thesis entitled:
Reasons and Value: Decision-Making Under Moral Uncertainty
written by Jay Geyer
has been approved for the Department of Philosophy

_____
Graham Oddie, chair

_____
Alastair Norcross, committee member

_____
Michael Tooley, committee member

Date: _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Geyer, Jay (PhD, Philosophy)

Reasons and Value: Decision-Making Under Moral Uncertainty

Thesis directed by Professor Graham Oddie

The problem of moral uncertainty can be described as the problem of what an agent ought to do when they are uncertain what the morally right thing to do is, and their uncertainty about how to act stems from uncertainty about moral facts salient to their decision. There are, broadly speaking, three responses to this problem. According to the first, agents ought to consider the moral tradeoffs of their prospective actions. That is, they ought to consider not only the likelihood of one action being better than others, but also how much morally better or worse it would be to choose one action rather than another, given their uncertainty. One way to do this is to maximize the expected moral value of one's actions. This is precisely the response to the problem of moral uncertainty that I refine and defend in this dissertation.

In the first paper, I present objections to an opposing view in the literature that rejects the entire project of giving a unique normative theory for moral uncertainty. I argue that two of the reasons cited in favor of this position by its proponents are in fact points of embarrassment for the position. In the second paper, I defend my response to the problem of moral uncertainty from an objection, according to which my response has counterintuitive implications for moral praise and blame-worthiness. I argue that this objection only succeeds against a very particular fine-grained version of my response, a version which no one in fact defends. Also in this paper, I use this objection as an opportunity to clarify and refine my position in the literature. In the third and final paper, I defend my position against two significant objections in the literature. I argue that both objections rely on a flawed way of modelling decision-making under moral

uncertainty. I offer an alternative, superior model. On this model, these two objections disappear. A competing theory to my own is thus in serious trouble as a primary motivation for this view was the seeming intractability of the one of the objections that I dispatch.

TABLE OF CONTENTS

INTRODUCTION AND OVERVIEW

Normative ethical theories describe our all-things-considered moral reasons to act in light of whatever the theories take to be relevant considerations, such as rights, consequences, duties, etc. We can qualify these all-things-considered moral reasons as objective or subjective depending on whether the reasons are sensitive to an agent's descriptive beliefs and evidence. For example, it is widely accepted that if an agent is justifiably certain about the relevant descriptive features of their choice, and if they know that they would have an all-things-considered moral reason to ϕ were those descriptive beliefs accurate, then they in fact have an all-things-considered *subjective* moral reason to ϕ. Even if it turns out that ϕ-ing is objectively morally wrong because the agent holds false descriptive beliefs, the agent would be morally blameless for ϕ-ing under these circumstances. Subjective moral reasons can thus be thought of as a second order set of moral norms based on an agent's beliefs, typically distinct from the norms furnished by the first order moral theories which take as input the considerations mentioned above.

These second order norms are fairly uncontroversial in the case of *certain* but false descriptive beliefs, and there is likewise little controversy about *uncertain* but false descriptive beliefs, at least in certain kinds of cases. Here is one such uncontroversial case: Suppose an agent's credence is 0.6 that the descriptive features of their situation are such that they have a very weak moral reason to ϕ, and 0.4 that the descriptive features are such that they have an extremely strong reason not to ϕ. The right thing to do in a case of descriptive uncertainty like this is to refrain from ϕ-ing. ϕ-ing under these circumstances would be reckless. One should hedge instead.

But there is a different kind of uncertainty for which it is quite controversial what one should do. The second order norms for this kind of uncertainty are in dispute, and it is even

disputed whether this kind of uncertainty warrants a unique set of second order norms at all. The uncertainty in question is moral uncertainty. Suppose an agent's credence is 0.6 that the *moral* features of their situation are such that they have a very weak moral reason to ϕ, and 0.4 that the *moral* features are such that they have an extremely strong reason not to ϕ. According to one position, the one that I will defend and refine in this dissertation, the morally uncertain agent is required to hedge just as the descriptively uncertain agent was. Call the family of theories that comprise this position 'Hedging Theories'.[1] According to a second position, the agent should not hedge. Call the family of theories that comprise this position 'Non-Hedging Theories'.[2] According to a third position, the agent's moral uncertainty does not factor into their subjective moral obligations, as it would if they were descriptively uncertain. The agent should do what the true moral theory would prescribe – perhaps that means ϕ-ing, perhaps not, depending on what ϕ-ing represents in this case. Call this theory 'Normative Externalism' or simply 'Externalism'.[3]

Normative Externalists reject the notion that we need a second order theory for moral uncertainty in the same we have for descriptive uncertainty. We can thus group Hedging and Non-Hedging Theories in a single camp in opposition to Normative Externalism. Call this collective camp 'Normative Internalism' or simply 'Internalism'.

Hedging Theories, because they straightforwardly co-opt the second order norms for descriptive uncertainty occupy a kind of default position in the literature. They inherit the plausibility of the second order normative approach to descriptive uncertainty, and place the initial burden of proof on their opponents to show why agents should not hedge in cases of moral

[1] Advocates of Hedging Theories include Graham Oddie (1994), Jacob Ross (2006), Alexander Guerrero (2007), and Andrew Sepielli (2008).
[2] Advocates of Non-Hedging Theories include Edward Gracely (1996), Johan Gustafsson and Olle Torpman (2014), and William MacAskill (2017). Ted Lockhart is plausibly grouped in this camp as well (2000). At least, his theory would prescribe not hedging in our present case.
[3] Advocates of Normative Externalism include Brian Weatherson (2014), Elizabeth Harman (2015), and Brian Hedden (2015).

uncertainty when they should in every other kind of uncertainty. After all, hedging is not only plausible for certain choices under descriptive uncertainty pertaining to *moral* norms, but also for choices under descriptive uncertainty pertaining to *prudential* norms. No one would invest money in a retirement account that has a 0.4 probability of being a Ponzi scheme, for a 0.6 chance of a modest return on one's investment.

Externalists and Non-Hedgers have responded by advancing three objections to Hedging Theories. They are: 1. the Problem of Intertheoretic Value Comparison, advanced by both Externalists and Non-Hedgers, 2. the Fetishism Problem, advanced by Externalists against both varieties of Internalism, and 3. the Exculpation Problem, also advanced by Externalists against both varieties of Internalism. In this dissertation, I argue that although 1 is decisive against Hedging Theories as they are typically modelled, there is an independently preferable way to model the problem of moral uncertainty on which 1 does not arise for Hedging Theories. I argue that 2 is actually a problem for *Externalism* and Non-Hedging Theories, but not for Hedging Theories. And I argue that 3 is actually uniquely a problem for *Externalism*, not the other way around. The upshot is that one of the Hedging Theories, in particular the one that directs agents to *maximize* their expected moral value, is the true second order normative theory for moral uncertainty.

In the first paper, I refute Externalism. I argue that two of the considerations offered in favor of the view are in fact embarrassing for it. These are the Fetishism Problem and the Exculpation Problem. According to Brian Weatherson, Externalism alone allows agents to be motivated by the proper objects of morality instead of fetishistically acting out of a desire to *do the right thing*, read *de dicto*.[4] I argue that Externalism entails that moral motivation is

[4] Weatherson (2014).

normatively irrelevant. Thus, considerations that take as their starting point the normative salience of virtuous or vicious moral motivation are embarrassing for the Externalist, but not necessarily for the Internalist. According to Elizabeth Harman, Externalism alone generates intuitively plausible results about praise and blame for agents who act objectively wrongly as a result of their false moral beliefs.[5] I argue that the cases Harman appeals to are faulty as intuition pumps, and that our moral intuitions regarding cases of objectively wrongful action stemming from false moral beliefs in fact count against Externalism.

In the second paper, I consider Harman's Exculpation Problem in more detail. I argue that it only succeeds in impugning Internalist theories that pair subjective probabilities (not epistemic probabilities) with moral prescriptions (not instrumentally rational prescriptions). Because no Internalist endorses that particular pairing of probabilities and reasons, the Exculpation Problem fails. I then consider whether Internalist theories should be understood as issuing moral or rational prescriptions, and offer some reason to favor the rational reading of Internalism. Moral prescriptions, I argue, would generate moral contradictions between the prescriptions of first order moral theories the agent is uncertain about (or better, the agent's direct moral judgments, as I argue in the third paper), and the second order prescriptions of Internalist theories. Rational prescriptions would not.

In the third paper, I tackle the Problem of Intertheoretic Value Comparison, and return to the Fetishism Problem. The Problem of Intertheoretic Value Comparison is widely viewed as decisive against Hedging Theories and is the primary motivation for Non-Hedging Theories.[6] In a nutshell, the problem states that Hedging Theories require non-arbitrary comparisons of moral value across separate and incomparable moral value functions each provided by the first order

[5] Harman (2015).
[6] See, for example, Gracely (1996), Gustafsson and Torpman (2014), Hedden (2015), and MacAskill (2017).

theories the agent's credence is divided among.[7] I argue that the problem rests on a bad way of modelling moral uncertainty. Moral uncertainty should be modelled as a bottom-up problem, with the agent's credence divided among their direct moral evaluations of their prospective actions, not as a top-down problem, with the agent's credence divided among first order moral theories each furnishing their own moral value function. On the bottom-up model, the agent can quite easily compare moral values across their divided moral evaluations because they are, after all, simply their own direct moral judgments. This solves the Problem of Intertheoretic Value Comparison, and it also solves the Fetishism Problem, at least for (maximizing) Hedging Theories. Non-Hedging Theories, I argue, are still on the hook for requiring agents to be moral fetishists. Moreover, with the Problem of Intertheoretic Value Comparison solved, there is simply nothing to recommend Non-Hedging Theories over Hedging Theories in light of the superiority of the bottom-up model that I advance.

The result of all of this is that Hedging Theories, on the bottom-up model, stand unrivalled as the leading answer to the question of what one should do when one's all-things-considered moral reasons seem to conflict. Agents ought to consider the expected moral value of their actions. In certain cases this means that they ought to hedge to avoid being morally reckless. This has important implications for a number of ethical puzzles in which the moral peril associated with the two choices is highly asymmetric – that is, where performing one action and being mistaken would be much worse than performing the alternative action and being mistaken. While I do not develop these implications in the dissertation, concerns about moral recklessness provide *prima facie* moral reasons to not eat meat, to not have an abortion, and to

[7] Jacob Ross (2006) and Brian Hedden (2015) each offer extensive treatment of this problem.

give altruistically to the point of marginal utility. These are just a few of the ethical implications of the arguments presented in this dissertation.

PAPER 1:

FETISHISM, BLAME, AND UNCERTAINTY:

WHY NORMATIVE EXTERNALISM'S PURPORTED STRENGTHS ARE IN FACT WEAKNESSES

I. INTRODUCTION

The literature on moral uncertainty can be divided into two camps – those who take an agent's moral beliefs (and possibly their moral evidence as well) to be relevant in determining how they ought to act under moral uncertainty, and those who do not. We may call the first camp 'Normative Internalists' or more simply, 'Internalists', and the second camp we may call 'Normative Externalists' or more simply, 'Externalists'. Normative Internalists disagree about how exactly an agent's moral beliefs inform their subjective reasons for acting. Consider, for example, the case of Victor, who is weighing whether to order veal or a vegan salad at a restaurant. Victor has a minor moral reason in favor of ordering the veal – he knows that, unlike ordering the salad, ordering veal would support local agriculture, which he knows to be a morally good, though relatively insignificant, thing. Victor also believes it to be slightly more likely than not that veal calves count for nothing morally, but he thinks it slightly less likely than not that they matter a great deal, such that ordering the veal would be seriously morally wrong, on a par with murder.

Some Internalists say that Victor ought to hedge under these conditions, choosing to not order the veal because doing so would be morally reckless. This might be because ordering the veal would have a very low expected moral value. That is, the probability-weighted moral

betterness or worseness of ordering the veal would be very low compared to ordering the salad.[8] Others say that Victor should go ahead and order the veal. This might be because, according to the moral theory Victor places the most credence in, ordering veal is the morally best action he can perform.[9]

But Normative Externalism stands in stark contrast to both of these views. According to Externalism, in the case as described it is underdetermined which action Victor should perform. This is because we do not know what Victor ought to do according to the *true* moral theory.[10] If in fact veal calves are morally valuable, then he should not order veal. And if in fact they are not (and he has some obligation to support local agriculture), then he should order the veal. But the Victor's moral uncertainty is entirely irrelevant to his subjective moral obligations. According to Normative Externalism, normative beliefs and evidence do not inform an agent's moral reasons in any way. As Brian Hedden puts it, 'there is no normatively interesting sense of *ought* in which what you ought to do depends on your uncertainty about normative facts.' Instead, he argues, an agent's subjective moral reasons are sensitive only to their uncertainty about descriptive facts – 'beliefs about descriptive matters make a difference to how you ought to act, while beliefs about normative matters do not.'[11] Here is how fellow Externalist Elizabeth Harman puts it:

> A person's moral beliefs and moral credences are usually irrelevant to how she (subjectively) should act. How a person (subjectively) should act usually depends solely

[8] Graham Oddie (1994), Jacob Ross (2006), Alexander Guerrero (2007), and Andrew Sepielli (2008) all endorse some sort of hedging view.
[9] This particular reason not to hedge is captured by the view known as 'My Favorite Theory', and is endorsed by Edward Gracely (1996), and Johan Gustafsson and Olle Torpman (2014). William MacAskill advances a different, related, theory that would also prescribe ordering the veal (2016). MacAskill's view frames moral uncertainty as a kind of voting problem, in which the different theories an agent is uncertain about act as voters. The primary motivation for these kinds of views is that they avoid having to make value comparisons across theories – a requirement for hedging that runs afoul of what is known as the 'problem of inter-theoretic value comparison'.
[10] Externalism is defended by Brian Weatherson (2014) and Elizabeth Harman (2015), though Harman calls the view 'Actualism'. Brian Hedden (2015) also endorses this view.
[11] Hedden (2015)

> on her non-moral beliefs and credences: her moral beliefs and credences are relevant only insofar as they provide warrant for beliefs and credences about what her non-moral situation may be.[12]

And finally, a third Externalist, Brian Weatherson:

> (Normative Externalism) is the view that the most important norms concerning the guidance and evaluation of action and belief are external to the agent being guided and evaluated. The agent simply may not know what the salient norms are, and indeed may have seriously false beliefs about them. The agent may not have any evidence that makes it reasonable to have true beliefs about what the salient norms are, and indeed may have misleading evidence about them. But this does not matter. What one should do, or should believe, in a particular situation is independent of what one thinks one should do or believe, and (in some key respects) of what one's evidence suggests one should do or believe.[13]

According to Normative Externalism then, an agent's subjective moral obligations are determined by their purely *descriptive* beliefs together with the true moral theory, however unknown or even epistemically inaccessible that theory is to the agent. Agents should act as the true moral theory prescribes, given their descriptive beliefs.

There are three arguments in the literature offered in favor of Normative Externalism. The first is presented by Elizabeth Harman. Harman argues that Externalism, unlike Internalism, yields the result that acting objectively wrongly due to moral ignorance is morally inculpatory, which she argues is intuitively the right result. Because Externalism better coheres with our praising and blaming intuitions, which are some indication of the true nature of our subjective moral reasons, Externalism is superior to Internalism.[14] The second argument is advanced by Brian Weatherson. Weatherson argues that Externalism, unlike Internalism, can accommodate the intuitive view that agents should be concerned with the things that matter morally, not with *doing the right thing*, where this is read *de dicto*. Caring about *doing the right thing* in this sense

[12] Harman (2015)
[13] Weatherson (2014)
[14] Harman (2015)

fetishizes morality in a way that is objectionable, and Externalism alone avoids this objectionable result. [15] The third argument in favor of Externalism is advanced by Brian Hedden, who argues that Normative Externalism, unlike hedging theories (but like non-hedging theories), avoids the problem of intertheoretic value comparison, which Hedden argues is intractable. The problem, to put it succinctly, is that hedging theories require agents to compare moral value across different first order moral theories, a procedure that cannot be executed without arbitrarily assigning moral values for those theories. [16]

Elsewhere, I defend Normative Internalism, and hedging theories in particular, against these objections.[17] Here, my aim is to hoist the Externalists on their own petard. I will not have anything to say about the problem of intertheoretic value comparison, because I do think Externalism succeeds in avoiding this problem.[18] But I argue that the first two considerations are in fact embarrassing to Externalism, not points in favor of the theory. Externalism gets exactly the wrong results when it comes to moral praise and blame-worthiness for morally uncertain or morally ignorant agents. I present several cases which tell strongly against the Externalist analysis, and I argue that the cases that Harman appeals to involve tainted intuitions. The situation is even worse for Externalists when it comes to concerns about fetishism. I argue that if Externalism is true, then an agent's moral motivation, fetishistic or otherwise, has no bearing on any normative assessment. Whereas Internalist views have ways of potentially responding to the

[15] Weatherson (2014)
[16] Hedden is the not the first to raise the issue of intertheoretic value comparison (see Ross 2006), but he is the first Externalist to wield the problem against hedging theories.
[17] See 'Moral Uncertainty and Moral Culpability' forthcoming in *Utilitas*, and 'Two Models of Moral Decision-Making' (unpublished ms).
[18] I argue that hedging theories also avoid this problem once we correct a mistaken way of framing moral uncertainty problems in a separate paper, 'Two Models of Moral Decision-Making'.

fetishism worry, Externalism, along with some very plausible additional assumptions, entails that concerns about moral motivation should be disregarded as normatively irrelevant.[19]

II. MORAL IGNORANCE AND CULPABILITY

No one in the moral uncertainty literature disputes that acting out of justified, but false, descriptive beliefs exculpates the agent who has done the objectively wrong thing. If a doctor, call her 'Diana', prescribes a drug that she is certain, and has every reason to be certain, will cure her patient, but due to some complication previously unknown to medical science, in fact causes the patient's death, she is not morally culpable, even though she is causally responsible, for the death. Diana did as she was subjectively morally obligated to do, according to every position in the literature. She acted as best she could have in light of her beliefs and best evidence about the *descriptive* facts relevant to her decision.

There is likewise unanimous agreement in the literature about the internalist justification for objectively wrongful behavior due to descriptive *uncertainty*. To borrow an example from Brian Weatherson, if someone, call her 'Carla', has baked a cake for a friend that she thinks is probably not poisoned, but just might be lethally poisoned, she should hedge and not serve the cake to her friend, even if refusing to do so is the objectively wrong thing – for example, if the cake is not in fact poisoned and her friend will be bitterly disappointed that she cannot eat it, etc.[20]

But now consider an agent who does the objectively wrong thing due to justified, but false, *moral* beliefs. Dolores is a doctor working in an emergency room who has just received an

[19] In 'Two Models of Moral Decision-Making' I argue that on the right model of moral decision-making worries about fetishism actually count in favor of hedging theories, but not non-hedging theories.
[20] Weatherson (2014)

unconscious patient who is a Jehovah's Witness and has lost a lot of blood. Dolores knows that the patient will die without an immediate blood transfusion, and also knows that because he is a devout Jehovah's Witness, the patient would never consent to a blood transfusion, even to save his life. Dolores is confident that she could administer the life-saving blood transfusion before the patient regains consciousness and before any of his friends and family arrive – no one would need to know how his life had been saved. But Dolores is certain that it would wrong to paternalistically override the patient's wishes about his own life, and her certainty is well founded – she has steeped herself in the bioethics and normative ethics literature and, through her diligent inquiry, has become completely convinced that it is wrong to paternalistically override an agent's autonomous desires about their own life. But suppose she is wrong about this. Suppose some form of consequentialism is the true moral theory, according to which the agent's overall interests, not the exercise of their autonomy, is what matters most morally, and suppose it is clear to Dolores that the patient's overall interests would be better promoted if she saved his life.

Internalists, of either the hedging or non-hedging variety, will say that Dolores is exculpated for much the same reason Diana is. She acted as best she could have in light of her best evidence about the *normative* facts relevant to her decision. But Externalists disagree. Given her descriptive beliefs (which in this case we may presume are accurate, but they needn't be on the Externalist analysis), the true moral theory would direct Dolores to save the patient's life. Thus, Dolores failed to do what she was *subjectively* morally obligated to do, and is thus culpable for her actions.[21] It is appropriate to blame Dolores for letting her patient die.

[21] What makes the obligation subjective for the Externalist is that it is sensitive to Dolores' descriptive beliefs. Had those been wrong – for example had Dolores falsely believed her patient to be an atheist who wanted the life-saving blood transfusion – then Dolores, who would have saved his life in that situation, would be morally blameless, even praiseworthy.

Harman argues that, in general at least, if not in this case, this is intuitively the right result – moral ignorance does not morally exculpate, and so the subjective moral prescriptions generated by Internalism in this kind of case must be wrong and Internalism must be false. Harman outlines her argumentative strategy as follows:

> To know whether false moral views exculpate, we must confront cases of false moral views head on, and ask whether they involve blameworthiness. When we do, we see that many cases of wrongful behavior by agents caught in the grip of false moral views are *paradigm* cases of blameworthiness.[22]

The first thing to note about this strategy is that to establish the blameworthiness of acting objectively wrongly due to moral ignorance requires more than finding *many* cases of intuitive blameworthiness. One would need to establish that in all or nearly all cases, these kinds of actions are blameworthy. This would require more than offering some favorable cases. One would also need to explain away the unfavorable ones. For example, I take the case of Dolores to be one that is unfavorable to the Externalist analysis. Intuitively, Dolores is not morally blameworthy for respecting the wishes of the Jehovah's Witness even if she did the objectively wrong thing. Harman owes us an explanation of why our intuitions tell against Externalism in this case, if her central argumentative strategy in favor of Externalism is that it yields intuitive results about praise and blame. If it turns out that there are roughly equal numbers of cases that tell in favor of the Externalist analysis and in favor of the Internalist analysis, then far from confirming either view, we should just abandon the strategy of motivating either view by pumping intuitions about praise and blame.

[22] Harman (2015)

I argue that we need not accept a stalemate on this kind of strategy. The cases that seem friendly to the Externalist analysis can be satisfyingly explained away, while the cases friendly to the Internalist analysis cannot. Intuitions about praise and blame for agents acting out of moral ignorance and uncertainty in fact strongly favor Internalism. Let us consider the cases that Harman takes to be 'paradigm cases of agents blameworthy for their wrongful actions':

> 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 is a gang member who believes that she has a moral obligation to kill a member of a neighboring 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.[23]

I share Harman's intuition that Max and Gail are blameworthy. But this intuitive result is in fact compatible with the Internalist analysis. As I argue elsewhere, Internalists can and do interpret their theories as either issuing rational, not moral prescriptions, or as formulating the agent's subjective reasons with respect to the agent's *justified* moral beliefs.[24] On either of these readings, Max and Gail's intuitive moral culpability is consistent with Internalism, either because their actions were rationally, but not morally justified, or because they have implicitly failed to satisfy the evidentiary requirements for holding justified moral beliefs.

I will briefly recount those arguments here, in order to show that the strongest cases in favor of Externalism do not in fact support the view – Internalism, properly understood, does just as well by our intuitions in these cases. After thus disarming Harman's cases in favor of

[23] Harman (2015)
[24] Geyer (2018)

Externalism, I will proceed to cases in which our praising and blaming intuitions are decidedly against Externalism.

Our culpability-finding intuitions about Max and Gail are only an embarrassment for a version of Internalism understood to issue *moral* prescriptions in light of an agent's *subjective* probability estimates. So one way to avoid this worry is to understand Internalism as issuing instrumentally rational prescriptions. Understood this way, there is no conflict between our intuitions that Max and Gail are morally culpable and Internalism's result that they do as they *rationally* ought. One can be perfectly rational and morally culpable. A number of Internalist authors clearly understand their theory this way, including Ted Lockhart, Jacob Ross, and Krister Bykvist.[25] Bykvist in particular holds this view precisely to avoid embarrassing implication such as the exculpation of Max and Gail.

The other way to avoid the counter-intuitive result that Max and Gail are exculpated is to understand Internalism as directing an agent to act in light of their *justified* moral beliefs, or their *epistemic* probabilities. Understood this way, it seems perfectly open to the Internalist to chalk up the intuitive culpability of Max and Gail to their seeming lack of epistemic justification for their false moral beliefs. This evidentialist interpretation of Internalism is explicitly endorsed by Andrew Sepielli, though other authors may also have this sort of view in mind.[26] This is compatible with the interpretation of Internalism as issuing rational prescriptions – this combined view would have Internalist theories issuing rational prescriptions in light of an agent's epistemic probabilities. Let us explore this response further by considering whether our culpability finding intuitions for Max and Gail are in fact due to their implicit lack of epistemic justification. If this is what explains our intuitions, then not only does the evidentialist version of Internalism avoid

[25] Lockhart (2000), Ross (2006), Bykvist (2014)
[26] Sepielli (2017)

any embarrassment caused by these cases, but the cases may in fact be embarrassing to Externalism. If *justified* moral ignorance does intuitively exculpate Max and Gail, then Externalism gets exactly the wrong results. I will attempt the more ambitious strategy of turning these cases against Externalism first. Even if I am unsuccessful, I will argue that it is because our culpability-finding intuitions are hopelessly tainted by features of the case. If this is so, then the cases of Max and Gail are at best neutral for Harman – they neither tell in favor of Externalism nor against it.

Harman anticipates that the Internalist will object that our culpability-finding intuitions are tracking a lack of epistemic justification for Max and Gail's false moral beliefs. Harman responds by claiming that 'it is a grave mistake to think that people cannot become convinced of deeply false moral views such as these' without violating some epistemic obligations. She then stipulates that Max and Gail have come by their beliefs by epistemically respectable means and that their false moral beliefs are justified.

I agree with Harman that agents may be justified in coming to hold false, pernicious moral beliefs, but adding this stipulation as an addendum to the cases of Max and Gail invalidates the initial intuitive response to the cases to the effect that Max and Gail are blameworthy. One cannot present the cases and then later stipulate that Max and Gail's beliefs are justified, and then lay claim to the same initial intuitions. One would have to build new cases from scratch that incorporate the justification of the agent's false moral beliefs in a more diligent way, especially because it would be *surprising* to find that someone could justifiably believe that mob executions and tit-for-tat gang killings are morally required. I will now attempt to do just that, although I expect I will not sway every reader's intuitions in favor of Max and Gail's exculpation. But if I fail, I argue that this will be due to intuition-tainting features intrinsic to the

case. Some readers, I expect, will simply never be able to embrace the notion that someone could be epistemically justified in believing that a mob hit or a tit-for tat gang killing is morally obligatory. Nevertheless, consider:

> Mitch 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. These beliefs spring from the way Mitch was raised and have been reinforced by all of the moral authorities in his community – a community in which his mafia family is quite popular. At times in his life Mitch has had genuine doubts about whether killing for the family is morally right, and he expressed these doubts to everyone in his life whose moral opinions he respects. But each of these moral authorities, people who display a great number of other moral virtues, confirm to Mitch that disloyalty to the family would be the greatest dishonor he could bring upon himself and his community. In fact, the only person in Mitch's life who questions this moral code is otherwise a vicious and untrustworthy person. On reflection, it is difficult for Mitch to believe that this person has the morally right views on this issue.

Is Mitch morally culpable for killing innocents out of a (false) sense of moral duty? I am inclined to say that he is not, and if the reader shares my intuitive response, then not only have we established that it was an implicit lack of epistemic justification that was tainting our intuitions in the cases of Max and Gail, but we've also appropriated these cases on behalf of the Internalist. Externalism gets exactly the wrong results even in cases that were selected to bolster the view, the case of Mitch being alike in all relevant respects to the case Max, except that Mitch is more plausibly epistemically justified in holding his false moral beliefs.

But even if the reader does not share my intuitions about Mitch, we can establish my second claim, that our intuitions are being tainted by an implicit lack of epistemic justification. Even though Mitch has a more robust epistemic backstory, it may be impossible for the reader to truly embrace the idea that his objectively heinous moral beliefs are justified. 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.

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.[27] 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.

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 in fact 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

[27] 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.

action). But unlike with 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 ϕ-ing is left unspecified, seems like the sort of thing that, *as far as we know* someone could be justifiably mistaken about. All of this indicates that the intuitions we have in the Max and Gail cases are tainted by their implicit lack of epistemic justification. Once we correct for this, the intuitive result in these cases is that the agent is morally blameless for acting objectively wrongly.

What these cases show is that intuitions generated by the cases Harman offers simply do not count as data against the claim that justified moral ignorance exculpates. At best, these cases are neutral for the Externalist. And it would be difficult if not impossible to offer other cases that would satisfactorily avoid the worry about tainted intuitions, while also intuitively impugning the version of Internalism that issues prescriptions in light of an agent's *justified* moral beliefs. Any attempt to argue from cases to the claim that Internalism entails that morally culpable agents are exculpated faces a dilemma. Either our culpability-finding intuitions will be strong, as in the cases of Max and Gail, 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 perverse implication. In the second case, they may simply shrug their shoulders and embrace the no-longer-embarrassing implications.

But if we think that Mitch is morally exculpated, then the cases of Max and Gail are in fact embarrassing for the Externalist. Glinda and Matt might be structurally similar, but Mitch *just is* the case of Max, but with a more good faith effort to do what Harman intended to do –

depicting Max as epistemically justified. To be sure, Externalism gets the right result for cases of unjustified moral ignorance, as the reader may have initially read into the cases of Max and Gail, but so does Internalism. And the fact that Internalism gets the cases of justified moral ignorance right too, suggests that it gives a better and more unified explanation of why epistemically unjustified agents are morally culpable for acting objectively wrongly due to their false moral beliefs. It is not because they happened to act contrary to the true moral theory, it is because they were in some way negligent in forming their moral beliefs.

So far we have only considered one kind of case. Dolores, Max, Gail, Mitch, Glinda and Matt were all agents who acted objectively wrongly on the bases of justified, but false, moral beliefs. If I am correct, then none of these agents are blameworthy, contrary to Externalism. Now let us shift to a different kind of case, ones in which the agent acts objectively rightly, but on the basis of unjustified, false moral beliefs. Consider the case of Connor:

> Connor has come to hold some poorly formed egoist moral beliefs. A friend told him that Ayn Rand was an egoist and Connor knows that Rand Paul was named after Ayn Rand. Connor takes Ron Paul to be a moral authority figure and so believes that egoism is true. Connor wants to murder his four year old cousin to steal his inheritance, and believes that, according to egoism, this means he is morally obligated to do so. While Connor is babysitting his cousin, he drowns him in the bathtub and frames it to look as though the cousin slipped and hit his head. Suppose the true moral theory is actual (not expected) utility act utilitarianism. Suppose further that by killing his young cousin, Connor unknowingly maximizes actual utility (fill in the details as you like).[28]

According to Externalism, Connor is morally praiseworthy. He has acted in compliance with the true moral theory, given his descriptive beliefs (which according to this moral theory are not relevant to what he ought to do). But clearly, Connor is not praiseworthy, even though he did the objectively right thing.

[28] This case obviously an intentionally borrows from James Rachels' famous Smith and Jones case.

Internalism, at least understood in the evidentialist way, gets the right result here.[29] Connor failed to have justified moral beliefs. Had he formed his beliefs responsibly, he surely would not have drowned his cousin. After all, the fact that doing so would maximize actual utility is, we may suppose, completely obscured from him. The Externalist might protest that there is no good reason to think that Connor couldn't have justified beliefs about egoism and thus come to the same conclusion. But this gets back to the dilemma I raised earlier. If egoism entails that Connor should kill his four year old cousin in order to claim his inheritance, then egoism is not a view anyone is epistemically justified in believing.

It is no use for the Externalist to say that Connor has not *really* acted in compliance with the true moral theory because he was not motivated by that theory. Externalism rules out an agent's moral beliefs as irrelevant to their moral reasons. And we may stipulate that there is nothing about this kind of utilitarianism that requires the agent to have a certain mindset while acting. It is a results-only view. Connor acts in compliance with the theory simply by getting the results it requires.

The Externalist might protest that I am saddling their theory with the undesirable implications of this particularly unappealing version of Utilitarianism. But Externalism cannot rule out this theory as being true. Externalism is a second order normative theory, not a partisan player at the level of first order moral theories. Moreover, the central issue is that Alex can act on any moral beliefs whatever, however poorly justified, and accidentally act in compliance with the true moral theory. But one is never morally praiseworthy for *accidentally* complying with the true moral theory. Consider Delilah:

---

[29] According to the rational prescription reading of Internalism (where the agent's obligations are grounded in their subjective, not epistemic probabilities), Connor would have done as he was rationally required, but that does not imply that he is morally exculpated.

> Like Dolores, Delilah is an emergency room doctor with a Jehovah's Witness patient who needs a blood transfusion to live, but who autonomously refuses that transfusion. Suppose the true moral theory would direct Delilah to let the patient die in these circumstances. Delilah lets the patient die, acting in compliance with the true moral theory. But Delilah does this because she is some sort of violent religious fanatic who falsely and unjustifiably believes that she is morally required to cleanse the world of Jehovah's Witnesses.

According to Externalism, Delilah has done the subjectively right thing. She has acted in compliance with the true moral theory in light of her descriptive beliefs, which we may stipulate are accurate and justified. We may stipulate also that the true moral theory places no moral restraints on the mindset or motivation of the agent – suppose it is the mere fact that the patient's autonomous request was granted, not that it was granted *because Delilah respects his autonomy*, that matters on this view. This means she is morally blameless, according to Externalism. But this is absurd. Delilah is morally monstrous even though she happened to do the objectively right thing.

All of this shows that, far from favoring Externalism, our intuitions about praise and blame in cases of moral uncertainty or moral false belief strongly favor Internalism. If these intuitions are any kind of reliable guide to the true nature of subjective moral obligations, then Externalism is false.

III. MORAL FETISHISM

The second major argument in favor of Externalism is raised by Brian Weatherson. Building on work by Michael Smith, Weatherson argues that Internalism requires agents to have fetishistic moral motivation. Externalism does not. Smith argues that morally virtuous agents are not motivated by a concern to do *what they believe is right*, where this is read *de dicto*. Being motivated by the goal of compliance with morality as such would 'alienate (them) from the ends at which morality properly aims' and would make them fetishists about morality. Alternatively,

virtuous, non-fetishistic agents care 'non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equality, and the like'. [30]

Weatherson agrees with Smith that caring about morality as such is a kind of vicious, or least less-than-wholly-virtuous, moral motivation:

> A good person will dive into a river to rescue a drowning child . . . and won't do so because it's the right thing to do. She'll do it because there's a child who needs to be rescued, and that child is valuable.[31]

Weatherson then argues that because Internalism directs the agent to act in light of their moral beliefs and evidence, they require an agent to act out of concern for morality as such. According to hedging theories, Victor should hedge, not because he cares about veal calves, but because ordering veal would be reckless, and therefore *the wrong thing to do*. Alternatively, if according to non-hedging theories Victor should not hedge, that too would be because he should be concerned with how he is behaving with respect to his moral beliefs – instead of hedging perhaps he should be acting in compliance with the moral theory he believes to be most probable. Either way, it is not the proper objects of morality that motivate Victor, but rather a concern about *getting it right*, morally.

Externalism avoids these worries, according to Weatherson. It does not direct the agent to ruminate over their own moral beliefs in an attempt to be in moral compliance. The agent's moral beliefs are entirely irrelevant to what they ought to do according to Externalism. Again, as Weatherson puts it, 'what one should do, or should believe, in a particular situation is independent of what one thinks one should do or believe, and (in some key respects) of what

[30] Smith (1994)
[31] Weatherson (2014)

one's evidence suggests one should do or believe.'[32] Agents are thus free to act for the right reasons, that is, out of concern for the proper objects of morality.

Elsewhere I argue that, once the right model of moral decision-making is adopted, fetishism concerns actually favor hedging theories. A virtuously motivated agent who is uncertain what to do will weigh both the magnitude of their moral concern for the proper objects of morality and their subjective probability that those concerns reflect moral reality. This means that Victor, if he really cares about the proper objects of morality, will hedge.[33] Here I argue that worries about moral motivation are in fact embarrassing for Externalism. This is because Externalism entails that moral motivation is morally irrelevant.

The problem can be glimpsed in the quote from Weatherson about the drowning child. A virtuous person will save the child not just because there is a child drowning, but also 'because that child is valuable'. This is right, of course. The mere fact that a child is drowning is not *by itself* a moral reason to save it. There must also be a normative component if there is to be a reason to act. The obvious candidate for the normative component in this case is that children are morally valuable – they create moral demands on us, for example to save them from drowning. Externalism can accommodate the fact that our reason to save the child has both a descriptive (the child is drowning), and a normative (the child is valuable) component. But it cannot accommodate that normative component as an *internalist* component. It cannot be because the agent *judges* or *believes* the child to be valuable that she has a reason to save it. It can only be because *in fact* the child is valuable. But, being virtuously morally motivated to ϕ requires the agent to evaluate that ϕ-ing is morally good. Here is the argument:

> P1. If Normative Externalism is true, then my moral evaluations never inform my moral reasons.

---

[32] Ibid

[33] 'Two Models of Moral Decision-Making' (Paper 3 of dissertation)

P2. If my moral evaluations never inform my moral reasons, then I cannot be virtuously motivated.
P3. But I can be virtuously motivated.
C. So, Normative Externalism is false.

P1 simply falls out of the definition of Normative Externalism. Moral beliefs in general, including moral judgments, etc., are irrelevant to one's moral reasons according to the view. P2 is making what I take to be a fairly uncontroversial claim about moral motivation. A necessary condition on being virtuously morally motivated to ϕ is that I evaluate ϕ-ing to be good. Motivational Internalists and Motivational Externalists will disagree about whether anything in addition to this moral judgment is required for an agent to be morally motivated, but all parties agree that the moral evaluation is necessary. P3 I also take to be obvious and is presupposed in Weatherson's argument against Normative Internalism. If P3 is false, then no one should care about the fetishism problem.

One way a Normative Externalist could try to resist the argument is to claim that there is an equivocation happening between the use of 'moral reasons' in P1 and P2. An Externalist might deny that moral evaluations give us any *genuine* moral reasons – that is, reasons that inform our subjective moral obligations – while allowing that our evaluations give us what we *take to be* moral reasons. Judging that the drowning child is valuable does not move the needle with respect to what we *really* have moral reason to do, according to Externalism. But on this reading, it does affect what we *take ourselves* to have moral reason to do and that is all that is required to *be virtuously motivated*. Virtuous moral motivation, on this response, is completely uncoupled from moral reasons. But it is still independently evaluable. Virtuously motivated agents care about the proper objects of morality, while moral fetishists display a kind of motivational viciousness. Internalist theories, to their discredit, still require agents to be

fetishists, because they require agents to *take themselves to have* the wrong moral reasons. Externalism, to its credit, makes no such requirements.

But why would merely taking oneself to have moral reasons of the right sort matter morally, if those reasons are not genuine? If both fetishistic and non-fetishistic moral motivation are equally irrelevant to what an agent subjectively ought to do, then surely it does not matter which kind of motivation they have. The issue is whether virtue or vice is a normatively relevant category if it is completely divorced from the rightness of action, as it would be on this account. I argue that it is not morally relevant. To be morally relevant, virtue would need to enter into some kind of explanatory or causal relation with right action. If you are a virtue ethicist, then think that virtue in some sense explains right action, and if you are not, then you think the order of explanation is the other way around. But on this Externalist response, an agent's having virtuous moral motivation neither explains the rightness of their action, nor is it explained by the (subjective) rightness of their action. Because it bears no such relations, it is morally irrelevant. And because Externalism entails that virtuous moral motivation is either impossible or morally irrelevant, considerations of virtuous motivation are in fact embarrassing to Externalism, not points in its favor.

Perhaps the Externalist could say that my non-genuine moral reasons (the ones I mistakenly take myself to have in light of my moral evaluations) have an *indirect* impact on how well I do morally. Being virtuously motivated by non-fetishistic moral judgments does not directly make my resulting actions right, nor does it explain the rightness of my actions in any way, but it may create a *propensity* for right action. Agents who act because they care about the proper objects of morality, even though this concern does not in any way determine or explain

the subjective rightness of their actions, may as a matter of fact tend to do the right thing more often.

But why should we think that this is true? If doing the subjectively right thing means acting in compliance with the true moral theory given our descriptive beliefs, wouldn't we hit that target more frequently if we employed some framework for rational choice? Plausibly, I will more frequently select the right action on the Externalist account if I act in accordance with the theory I take to be right, as certain non-hedging theorists prescribe, than if I care purely about the proper objects of morality. Our ground level judgments about the objects of morality often come into conflict, as the length and breadth of the normative ethics literature bears witness. Unless the true moral theory is a kind of particularism, then being non-fetishistic about moral motivation seems unlikely to result in doing the right thing at a greater frequency than being fetishistic, on the Externalist analysis.

So, Externalism entails that an agent cannot be virtuously morally motivated. Or at least, that if they are, that their virtuous motivation is morally inert. Far from being a point in favor of Externalism, considerations about virtuous moral motivation are incompatible with the view.

CONCLUSION

I have argued that two of the considerations offered on behalf Normative Externalism in fact are embarrassing for the view. Externalism gets intuitively exactly the wrong results with respect to praise or blameworthiness of an agent who acts out of false moral beliefs. Internalism can satisfactorily accommodate our intuitions in these cases. Externalism also fails to leverage the charge of fetishistic moral motivation against the Internalism. Instead being virtuously motivated, because it presupposes that our moral judgments give us moral reasons, is impossible

on the Externalist account. In light of these considerations Normative Externalism should be rejected.

PAPER 2:

MORAL UNCERTAINTY AND MORAL CULPABILITY[34]

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 favor 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.[35] Another family of theories recommends ordering the veal, because moral hedging requires the agent to make inter-theoretic value comparisons, a task that they cannot non-arbitrarily complete.[36] An example of such a non-hedging theory would be My Favorite

---

[34] Originally published in *Utilitas* (2018)

[35] Advocates of EMVT include Jacob Ross (2006), and Andrew Sepielli (2009). Advocates of hedging more broadly construed include Graham Oddie (1995) and Alexander Guerrero (2007).
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 MacAskill (2016). For purely aesthetic reasons, I opt for 'moral value' instead of the more cumbersome 'moral choice-worthiness'.

[36] 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

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.[37]

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.[38] 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 leveled 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.[39] Weatherson argues that this kind of moral motivation is fetishistic and thus an embarrassment for Internalism. In this paper, I focus on a second argument against Internalism, raised by Harman, which I will refer to as the 'exculpation problem'.[40] 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

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.

[37] MFT is advocated by Edward Gracely (1996), and Johan Gustafsson and Olle Torpeman (2014). William MacAskill also offers a non-hedging approach to moral uncertainty (2016).

[38] Brian Weatherson (2014), Elizabeth Harman (2015), and Brian Hedden (2015) all endorse this view.

[39] Weatherson (2014)

[40] Harman (2015)

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.[41] 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

[41] Authors have in fact independently advanced each of these responses. Andrew Sepielli (2017) responds to the exculpation problem by appealing to epistemic probabilities, while Krister Bykvist (2014) 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 author's arguments share certain core similarities to my response to Harman, I offer some important supplements and criticisms of their work. My paper 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.

objectively wrong.[42] 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 ϕ on the basis of their moral beliefs, then Internalism entails that moral ignorance exculpates.

Harman then argues that moral ignorance does not exculpate.[43] 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 neighboring 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.[44]

[42] 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 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.

[43] Harman argues for this claim first in an earlier paper (2011).

[44] Harman (2015)

According to Harman, Max and Gail are both 'paradigm cases of agents blameworthy for their wrongful actions'.[45] Harman argues that even if we fill in Max 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.

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*

[45] Ibid

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.[46]

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.

According to 2, a murderer who acted prudently (say the murder was in her interests and she was careful to not get caught) is thus not morally culpable for murdering. But this is absurd. [47] 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.

---

[46] 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' (2015). 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.

[47] 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.

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.[48] Similarly, Johan Gustafsson and Olle Torpman, non-hedging proponents, frame their question as which actions would amount to a 'morally conscientious choice' for a morally uncertain agent.[49] 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.[50]

Likewise, Ted Lockhart is quite clear throughout his book on the subject that his normative theory for moral uncertainty places rational, not moral, constraints on morally uncertain agents.[51] 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

[48] Oddie (1994)
[49] Gustafsson and Torpman (2014)
[50] Ross (2006) – my italics
[51] Lockhart (2000)

an intention that . . . maximizes expected choiceworthiness' and argues that this principle governs decision making under moral uncertainty as well as non-moral uncertainty.[52]

Krister Bykvist similarly argues that uncertain agents have *rational* reasons to act in light of their uncertainty, but not moral reasons.[53] 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.[54] Harman supposes that Internalists will still want to say that Bill should act on those beliefs in an appropriate way. For example,

[52] Wedgwood (2013) – my italics
[53] Bykvist (2014)
[54] Harman (2015)

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 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 judgment can be made sensible on the rationality reading. A plausible *corollary* to the rational reading of Internalism, though not necessarily a part of 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 paper, 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 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'.[55] 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 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 our culpability-finding intuitions in these cases are entirely due to their implicit lack of epistemic justification.

[55] Sepielli (2017)

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 likely 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. [56]

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.[57] Is Glinda morally culpable for acting out of her

[56] 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 favor of my claim that our intuitions are tracking a lack of epistemic justification.
Each of these strategies reinforce each other and underscore the flaws in Harman's original intuition pumps.

[57] 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.

false moral beliefs by having an abortion? I am strongly inclined to say that she is not, and I hope the reader will agree.

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 with 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.[58] 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

[58] The failure of these cases extends beyond Harman's 2015 paper against EMVT, to her earlier paper, 'Does Moral Ignorance Morally Exculpate?' (2011). In this paper, Harman is responding to skeptical worries raised by Gideon Rosen (2004) about to what extent 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.

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 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 either 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 the agent's beliefs are split among. 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 fetuses 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).[59]

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

[59] The inspiration for this case is obviously Judith Thomson's famous paper 'A Defense of Abortion' (1971).

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 it. 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 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 favor of first order theories that have similar decision-guiding principles. Because the rational reading of Internalism's prescriptions avoids this implication, it should be favored.

VI. OBJECTIONS

I will briefly consider two objections to what I have argued in this paper. 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.[60] The response Harman considers also involves the charge of equivocating on the sense of 'ought' and 'culpable' in the conditional 'If one ought 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 to only 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

[60] Harman (2015)

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 favored by non-hedging theorists allows them to skirt the problem of intertheoretic value comparison, thought to be a significant obstacle for hedging theories like EMVT.[61] 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 intertheoretic 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.[62]

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

[61] See e.g. Gustafsson and Torpman (2014) and MacAskill (2017)

[62] 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.

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 her 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 that Internalism is offering a normative theory for.

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 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.[63]

[63] I borrow this case from Weatherson (2014).

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.

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 successfully navigate the deontic conflict problem.

Some daunting problems remain for Internalism. For hedging theories, there is the problem of intertheoretic value comparison.[64] For Internalism more generally, there is the charge that the theory requires agents to be moral fetishists.[65] 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.

[64] See e.g. Ross (2006), and Brian Hedden, (2015).
[65] Weatherson (2014)

PAPER 3:

TWO MODELS OF MORAL DECISION-MAKING

I. INTRODUCTION

According to a leading view in the moral uncertainty literature, there are times when a morally uncertain agent is required to hedge their bets morally, refraining from acting in accordance with the moral theory they believe is most probably true, because the moral peril associated with that action is too great. Call any theory that would prescribe hedging under these kinds of circumstances a 'hedging theory'.[66] To illustrate, consider the case of Victor, who is weighing whether to order veal or a vegan salad at a restaurant. Victor has a minor moral reason in favor of ordering the veal – he knows that, unlike the salad, it would support local agriculture, which he knows to be a morally good thing. Victor also believes it to be slightly more likely than not that common sense morality is true, according to which veal calves count for nothing morally, but he thinks it slightly less likely than not that a Tom Regan-style animal rights theory is true, according to which veal calves matter a great deal, such that ordering the veal would be morally on a par with murder.[67]

Under these conditions, hedging theories would prescribe ordering salad, despite the fact that the majority of Victor's credence favors ordering veal. Ordering veal under these conditions would be morally reckless, akin to serving someone a cake that you are only slightly more confident than not is poison free. Expected Moral Value Theory (EMVT) is an example of a hedging theory. EMVT directs agents to maximize expected moral value, where 'moral value'

[66] Hedging theories are endorsed by Graham Oddie (1994), Jacob Ross (2006), Alexander Guerrero (2007), and Andrew Sepielli (2008).
[67] Regan (1987)

refers to the quantified relative moral betterness or worseness of Victor's prospective actions. The (subjective) probability weighted moral value of ordering salad is higher than the probability weighted moral value of ordering veal, so Victor ought to order salad, according to EMVT.

Hedging theories contend with two opposing views in the moral uncertainty literature. One of these views is similar to hedging theories in that they are both *internalist* views. By that I mean that according to both views, internalist factors such as an agent's moral beliefs and evidence inform what that agent subjectively ought to do. Both internalist approaches advance second order normative theories for choices under moral uncertainty in which agents' (perhaps justified) moral beliefs determine their subjective moral reasons. I will refer to the internalist alternative to hedging theories as 'non-hedging theories'. My Favorite Theory (MFT), endorsed by Edward Gracely, and Johan Gustafsson and Olle Torpman is an example of a non-hedging theory.[68] MFT directs an agent to act in accordance with the theory they have the highest credence in. Also included in the non-hedging category is William MacAskill's probability-weighted Borda Rule approach, which frames moral uncertainty as a kind of voting problem, with the moral theories the agent is uncertain about playing the role of voters.[69] As my naming scheme suggests, neither of these theories prescribe hedging – according to each of these theories, Victor should order the veal.

The third position in the moral uncertainty literature rejects the entire project of giving a second-order normative theory for moral uncertainty. According to this theory, which I will refer to as 'Normative Externalism', or simply 'Externalism', internalist factors like moral

[68] Gracely (1996) Gustafsson and Torpman (2014)
[69] MacAskill (2016). MacAskill's probability-weighted Borda Rule approach is meant for decision scenarios in which, according to MacAskill, the morally uncertain agent is unable to maximize expected moral value. For scenarios that are amenable to the maximizing approach, MacAskill's approach is more similar to hedging theories.

beliefs and evidence are entirely irrelevant to what an agent ought to do. A morally uncertain agent should act in accordance with the *true* moral theory given their descriptive beliefs, their uncertainty about, or even complete lack of epistemic access to, that theory notwithstanding. Brian Weatherson defends this theory, as does Elizabeth Harman, who refers to the theory as 'Actualism'. Brian Hedden likewise expresses support for this view.[70]

I argue that there has been a subtle, but ubiquitous and significant error made in the literature on moral uncertainty to date. The error is to model the decision problem facing a morally uncertain agent as a top-down problem. By that I mean that agents' moral uncertainty has been described as ranging over first order moral theories, each with their own agent-independent value function ranging over the agent's prospective actions. On this model, all morally evaluative aspects of the agent's decision originate in these first order theories. The agent's own beliefs are salient only insofar as they provide a probability distribution over those first order theories. I argue that this is a bad way to model choices under moral uncertainty. Because it only applies in cases in which the agent's credences are divided among theories that have agent-independent value functions, it needlessly limits the scope of the theory to a small subset of cases of moral uncertainty. Even worse, it misidentifies what is important in moral decision-making. For even when a decision under moral uncertainty exhibits the features necessary for the model to apply, those features are not in fact what informs an agent's subjective moral reasons. Instead, decisions under moral uncertainty should be modeled from the bottom up. By this I mean that the agent's uncertainty is best viewed not as ranging over first order moral theories, but rather as ranging over the agent's own direct, ground level moral evaluations of their actions. Once this model is adopted, two of the most significant objections

[70] Weatherson (2014), Harman (2015), Hedden (2015)

to hedging theories in the literature – the problem of inter-theoretic value comparison, and the fetishism problem – cease to be threatening. In fact, on the bottom-up model the fetishism problem becomes a consideration in favor of hedging theories.

In the first half of the paper, I describe the top-down model and distinguish it from the bottom-up alternative. Here I also introduce the problem of inter-theoretic value comparison as it neatly illustrates the top-down conception of moral uncertainty, and then show how this problem, widely viewed as decisive against hedging theories, fails to gain any traction on the bottom-up model. I then argue that the top-down model has significant flaws that the bottom-up model avoids. These provide independent grounds for preferring the bottom-up model. In the second half of the paper I develop the idea that fetishism concerns are considerations in favor of the bottom-up understanding of hedging theories, arguing that non-hedging theories cannot similarly avail themselves of the bottom-up model to avoid the fetishism issue. Moreover, with the problem of inter-theoretic value comparison neutralized, a problem which provides the main motivation for non-hedging theories, there is simply nothing to recommend this view.

II. TOP-DOWN VERSUS BOTTOM-UP AND ITS IMPLICATIONS FOR VALUE COMPARISON

The standard way of modelling Victor's moral uncertainty is to have Victor's subjective probability function range over the two moral theories he considers to be possibly true – common sense morality and animal rights theory. Each of these theories has a moral value function ranging over Victor's two prospective actions – ordering veal or ordering salad. The expected moral value of each of Victor's actions is determined by the cardinal rankings of moral betterness and worseness that results from each theory's moral value function, weighted by

Victor's subjective probability that the theory is true. I call this model 'top-down' because all of the moral value in the model originates in the first order moral theories that Victor is uncertain about. The value flows down, so to speak, from the theories themselves, whose value functions are independent of Victor's direct moral evaluation of his prospective actions. On this model, all Victor must bring to the table are his beliefs about the truth of those theories. On the top-down model, Victor is simply trying to comply with the true moral theory, but is uncertain which one it is. The value functions are Victor's value functions only derivatively insofar as his probability distribution ranges over them.

The alternative is the bottom-up model. On this model, it is Victor's direct moral evaluations of his actions that furnish the evaluative component of the decision scenario. On the bottom-up model, the expected moral value of ordering veal is *not* the value according to common sense morality weighted by the probability that this theory is true, plus the value according to animal rights theory weighted by the probability that this theory is true. Instead it is the moral value of ordering veal according to Victor's judgment that veal calves are morally worthless weighted by his subjective probability that this judgment is accurate, plus the value of ordering veal according to Victor's judgment that veal calves are very morally valuable weighted by the probability that this judgment is accurate. These judgments may correlate with, and flow from, certain first order moral theories, and Victor may even attribute them to his beliefs about the first order moral theories, but it is Victor's direct moral evaluations themselves, not the theories' value functions, that are the source of moral value in this model. The *actual* value functions of the first order moral theories play no direct role in informing Victor's subjective moral obligations, though these theories may have played a causal role in Victor's moral evaluations being what they are.

We can get clearer on the distinction between the top-down and the bottom-up model by way of an analogy. Suppose that instead of two moral theories, Victor has two moral *advisors* who give him conflicting advice about ordering veal. According to the top-down model, all the input that Victor provides that informs his subjective moral reasons with respect to ordering veal is his probability distribution over which of his advisors is most credible. It is the advisors' moral judgment that matters and Victor's role is simply to judge which advisor is most credible. On the bottom-up model, the advisors play at most an indirect role in determining Victor's subjective moral reasons. The advisors may inform Victor's moral evaluations about eating veal because he finds them persuasive, and Victor may believe one advisor is most probably right, but Victor's probability distribution is over his own, direct, divided (because he is morally uncertain) moral evaluations of ordering veal and ordering a vegan salad.

The reader may be wondering whether this construal of the top-down model is fair with respect to how divorced Victor's direct moral evaluations are from the value functions of the theories. Perhaps on the top-down model the theories' value functions are just a proxy for Victor's own direct moral evaluation of his prospective actions. Not so. On the top-down model, which is either explicitly endorsed or else implicitly accepted by most authors working in the literature, the theories' value functions are not simply a gloss for Victor's own direct moral evaluations of his prospective actions. They are independent of Victor's direct moral evaluations.

To see this, consider first what various authors have said about the structure of decisions under moral uncertainty. Brian Hedden writes that a hedging theory like EMVT 'evaluates each action by looking at how (subjectively) good (or bad) an action would be *according to each moral theory you take seriously* and discounting that goodness by your degree of belief that that

moral theory is correct.'[71] Brian Weatherson draws an analogy between moral uncertainty and prudential uncertainty in which the agents' own prudential preferences conflict with named welfare theories they have credence in. Acting contrary to their own preferences by engaging in activities that leave them cold in order to avoid prudential recklessness is supposed to be embarrassingly analogous to the behavior prescribed by hedging theories.[72] While several authors stress that the agent's credences may be distributed over something other than theories *per se* (Sepielli prefers 'normative propositions'[73]), whatever it is that are the objects of the agent's divided belief are widely understood to be the source of evaluative information in the decision scenario, not the agent's own direct moral evaluations.

Consider second that it would be impossible to make sense of the problem of inte-theoretic value comparison unless the theories' value functions were divorced from Victor's own direct moral evaluations, as they are in the top-down model. The problem of inter-theoretic value comparison is that each first order moral theory an agent is uncertain about has a moral value function that is entirely independent from and incomparable to, the value functions of every other theory. As Jacob Ross puts it, the problem is that 'one cannot compare value intervals *across ethical theories*.'[74] To illustrate, suppose that according to Animal Rights Theory it is 1000 times worse to order veal than salad, while according to Common Sense Morality, it is twice as bad to order salad rather than veal. These cardinal rankings are consistent with an infinite number of discrete moral value assignments. Perhaps the values of ordering veal compared to salad in Animal Rights Theory is -1000 and -1, while the values according to Common Sense Morality are -.5 and -1. Or, perhaps it is -1 and -.001 according to Animal

[71] Hedden (2015), my italics.
[72] Weatherson (2014)
[73] Sepielli (2008)
[74] Ross (2006), my italics.

Rights Theory and -50 and -100 according to Common Sense Morality. Either value assignment is consistent with the cardinal rankings provided by the two theories, but according to the first set of assignments eating salad has a higher expected moral value, while according to the second ordering veal does. Any choice among these or any other infinitely many sets of compatible value assignments must be arbitrary, making hedging theories' prescriptions arbitrary.

What hedging theories require but lack, according this objection, is a non-arbitrary way to bridge the value functions of the two or more theories the agent's credence is divided among. This would allow the translation of one cardinal ranking to another, allowing the agent to aggregate the rankings into a single cardinal ranking spanning all of the act-theory pairs of outcomes. But there is nothing in any of the first order theories that allows one to do this. Each theory is formulated under the assumption that it is the true moral theory. They do not offer a mechanism for converting their moral currency, so to speak, to that of another theory.

It is impossible to make sense of this problem unless the value functions of the first order moral theories are divorced from the agent's own moral evaluation. If the theories' value functions were just proxies for the agent's own evaluations, then there would be no issue with comparing them. In fact, there simply would not be two or more value functions to compare. On the bottom-up model, there would only be the agent's own, single value function derived from his divided moral evaluations of his prospective actions. The ubiquitous acceptance of the problem of intertheoretic value comparison as a daunting, if not insurmountable, objection to hedging theories is testament to the ubiquitous acceptance of the top-down model. On the bottom-up model, the problem simply does not arise – an agent does not need to compare value *between theories* on this model. They simply morally evaluate their prospective actions.

To put the difference between the top-down and the bottom-up model in sharp contrast, consider an agent whose direct moral judgment differs from the value function of the theory she takes to be true. Imagine a committed Act Utilitarian, call her Alexandra, who is a doctor, and is deliberating between killing a wandering loner and redistributing his organs to save two other patients, or letting the wandering loner live and the two patients die. Alexandra reports that she is certain that Act Utilitarianism is the true moral theory and acknowledges, at least in the abstract, that it would direct her to kill one to save two. But Alexandra feels an overwhelming sense of moral revulsion at the prospect of doing so in her present context. It is not mere squeamishness that Alexandra feels at the prospect of killing her patient to save two others, but an inescapable conviction that doing so would be very wrong. Nevertheless, Alexandra insists that act utilitarianism is the true moral theory. To resolve her contradictory beliefs, she tells herself some story about how letting the two patients die and the one live would in fact maximize utility (in fact it would not, as Alexandra should recognize). Because Alexandra's credence is 1 *that act utilitarianism is the true moral theory*, the top-down model would use the actual value function of act utilitarianism to furnish the moral values in this case, in which case killing the one to save the two is better than letting two die and one live. The bottom-up model would use Alexandra's direct, ground-level moral evaluation that it would be worse to kill one to save two than letting two die and one live.

It may seem that I am being unfair to the top-down model in the case of Alexandra. Could not the top-down model say that the proper value function is the one selected by Alexandra's direct moral evaluation? In that case, it is not the actual value function of Act Utilitarianism that informs her moral reasons, but rather some different value function that assigns higher moral value to letting two die and one live than to killing one to save two, at least

in her present act context. That is the value function that her credences are *really* distributed over, not the one that she *would report* her credences to be distributed over.

But I am not being unfair. The distinction between Alexandra's reported credence in the value function of Act Utilitarianism, and the value function generated by her direct moral evaluation of her actions precisely illuminates the distinction between top-down and bottom-up models of moral decision-making. If the top-down model used Alexandra's direct moral evaluation to generate the relevant value function instead of using the value function of the theory that she claims to have complete credence in, it would no longer be the top-down model, but the bottom-up model. We can see this by considering that the problem of inter-theoretic value comparison would not arise on this understanding of the origins of the value function. Suppose that, contrary to the original case, Alexandra is not certain of the truth of Act Utilitarianism, but is instead uncertain between that theory and another first order moral theory. If *the value function of Act Utilitarianism* is determined by Alexandra's direct moral evaluations that she happens to attribute, perhaps falsely, to Act Utilitarianism itself, then there is no issue with comparing the moral values of Act Utilitarianism and the other first order moral theory. After all, both theories' value functions are really just the product of Alexandra's direct moral evaluation, which she happens to attribute to the theories. Alexandra can easily compare value 'across the theories' because this just means reflecting on her own divided moral judgment, which constitutes a single value function. If the agent's direct moral evaluations determine the moral values of actions on a model, then the model is bottom-up, not top-down model.

These considerations in the case of Alexandra are the foundation of the first *independent* reason to favor the bottom-up model over the top-down model. (The model's easy avoidance of the problem of inter-theoretic value comparison would be another reason to favor it, if only for

hedging theorists.) In cases in which the agent's direct moral judgment conflicts with the value function of the first order moral theory they ostensibly believe in, it is most plausibly the direct moral judgment that informs their subjective moral reasons. The actual value function of Act Utilitarianism is irrelevant to what Alexandra subjectively ought to do. Even though Alexandra takes herself to be an orthodox Act Utilitarian, her actual moral preferences reveal *her* moral value function. And it is this value function, not the one she erroneously ascribes to herself that informs her subjective moral reasons. This is plausible on its face – the more reliable guide to what someone values is their actual preferences, which are often revealed by their choices, not what they profess to value. And, all else equal, it is an agent's actual moral evaluations, not their professed ones, that determine their subjective moral obligations. It is at the ground-level, so to speak, of direct action-evaluation, not at the theory-level, that subjective moral reasons are formed.

Here is another way to motivate this point. Suppose that instead of reconciling her cognitive dissonance by telling herself that letting the two die and the one live would in fact maximize utility, Alexandra recognized that her strong moral aversion to killing the wandering loner was in tension with the moral theory she professed to believe. Suppose that, in full awareness of this epistemic tension, Alexandra chose to maintain her orthodoxy to Act Utilitarianism, killing the one to save the two. But suppose she was never able to completely expunge her strong, negative, direct moral evaluation of killing her patient. She would later attempt to dismiss her moral horror at her own actions and the later pangs of remorse that she felt as merely an aversion to getting her hands dirty, but her direct belief that what she did was wrong would persist.

To clarify, Alexandra's situation was not one in which she was morally uncertain between two different moral theories each with their own agent-independent moral value function. Nor did she have conflicting direct moral evaluations. Her direct moral evaluation was undividedly that killing the wandering loner to save two others was wrong. But her credence about which first order moral theory is true was entirely in favor of Act Utilitarianism.

Is Alexandra blameworthy for killing her patient to save two others? Or at least, is she morally criticizable? I am inclined to say that she is, and perhaps surprisingly, I think that she is for reasons that are usually thought to count against hedging theorists – I think that in choosing compliance with Act Utilitarianism over her own direct moral judgments, Alexandra has fetishized morality. What really matters morally, and what virtuous agents care about, are what Michael Smith calls 'the proper objects of morality', things like the suffering or happiness of other people, justice being done, etc.[75] Improper objects of morality by contrast include compliance with morality for the sake of compliance. This is what Smith calls 'moral fetishism' – caring about *doing the right thing* read *de dicto*. Brian Weatherson takes these considerations to constitute an objection to hedging theories (if he is right, they would also constitute an objection to non-hedging theories), because these theories require the agent to care about moral compliance as such.[76]

Later in this paper I will address this objection in more detail. But here, I simply want to note that these considerations seem count against Alexandra acting in compliance with Act Utilitarianism given her contradictory direct moral evaluation of her actions. If Alexandra chose to act on her direct moral evaluations, contrary to Act Utilitarianism, she might very well be acting for virtuous reasons – out of concern for the wandering loner, say. But if she acts in

[75] Smith (1994)
[76] Weatherson (2014)

accordance with Act Utilitarianism, in the teeth of her own strong direct moral judgment, she could only be doing so out of a desire to comply with what she takes to be the true moral theory. Acting for such reasons would fetishize morality. This gives us reason to think that the bottom-up model is locating the source of moral evaluation in the more morally salient place – the agent's direct evaluations, not the theories their credence is divided among.

Another reason to favor the bottom-up model is that, unlike the top-down model, it applies to nearly every case of moral uncertainty. The only cases to which the top-down model apply are those in which an agent's credence is divided among first order moral theories that issue agent-independent value functions. But these constitute a relatively small subset of all moral uncertainty cases. Most people have never encountered or seriously considered any substantive first order theory, other than perhaps the Golden Rule. When they encounter moral uncertainty it is not because their credence is divided among two or more external value functions. It is because they have conflicting direct moral judgment.

For example, consider the case of Alfred. Alfred's wife has just gotten a new haircut. Alfred thinks the haircut makes his wife look much older than she really is. When his wife asks him how she looks, Alfred must decide whether to tell her the unpleasant truth or a pleasant falsehood. Alfred judges that both actions have a morally unsavory aspect – either hurting his wife's feelings or lying to her. But he also judges that hurting her feelings would be worse. Alfred's moral evaluations are simply the result of his direct moral appraisal of his prospective actions. Alfred has never heard of terms like 'consequentialism', 'deontology', 'categorical imperative', 'utility', and the like. If he reflected on it, perhaps he would agree that in general, it is wrong to be dishonest with your spouse, and also that in general it is wrong to hurt your spouse's feelings when it could be easily avoided. These might constitute the seeds of two

skeletal first order moral theories that vaguely resemble the divide between deontology and consequentialism. But is clearly a strained and absurd analysis of the case to say that Alfred is uncertain between these two first order moral theories, and must find a principled way to arbitrate their (incompatible) value functions. Alfred is morally uncertain. He has conflicting moral judgments and does not know what he morally ought to do. But he is not uncertain about *which moral theory is true*, in any ordinary sense of 'moral theory'. That does not enter his mind, and it would be bizarre and disingenuous to model his decision in a top-down way.

Because of this, the top-down model simply does not apply to a case like Alfred's, because there is no salient value function other than Alfred's own two direct moral appraisals of the situation. Of course, the top-down model could always be amended such that in these situations we may use the agent's direct moral evaluations to generate a single value function, but in situations where there are agent-independent, theory-generated value functions, we should use those instead. But what could motivate this bottom-up addendum to the top-down theory? Surely if the bottom-up model is good enough for cases like Alfred's, it is good enough for cases like Alexandra's, absent some compelling argument to favor using the theories' value functions in cases like Alexandra's.

I think is plausible that most cases of moral uncertainty are like Alfred's, not like Alexandra's. The overwhelming majority of moral agents are not ethically trained, and even most ethicists I suspect, typically resolve their moral uncertainty in a fashion more like Alfred's – not with respect to what they guess the value functions are of the theories their credences are distributed over, but by more direct, theory-independent means. Perhaps they are wrong to do so. Perhaps the ethically trained, who know better, should operate under the top-down model, assigning moral value to their actions only derivatively on the basis of which agent-independent

value functions are live possibilities by their lights. But the concerns raised over the case of Alexandra should make us question whether the ethically trained *must* arbitrate their moral uncertainty this way, even if in fact some do. If very few agents ever arbitrate their moral uncertainty in the way indicated in the top-down model, then the model risks being largely inapplicable as a framework for either a decision guide or an evaluative theory for moral uncertainty, which would give us good reason to reject it.

The bottom-up model, on the other hand, applies to nearly every case of moral uncertainty. Nearly every time an agent is morally uncertain, they have in their possession some direct moral evaluation of the moral betterness or worseness of their prospective actions. The only exception would be someone who really makes no moral evaluations outside of their probabilistic judgment about the truth of various first order theories. Such an agent would have no evaluative judgments about killing this patient or letting these patients die as such. They would only have judgments about the probability of Act Utilitarianism or some other theory, and have moral judgments about the patients only in a derivative, rather alien way. Such an agent could not resolve their moral uncertainty on the bottom-up model, but only on the top-down model.

But the subset of cases to which the bottom-up model would not apply is surely much smaller than the subset of cases to which the top-down model would not apply. It would be a highly unusual moral agent who had beliefs about what the value functions of Act Utilitarianism or Kantian Ethics would deliver with respect to some action, but who lacked any direct moral appraisal of that action. Even when one is a committed adherent to some first order moral theory, that commitment generates direct moral evaluations of actions. Agents who have really internalized Kantian Ethics tend to directly negatively evaluate particular actions that disrespect

other people's autonomy. They do not make such evaluations derivatively with reference to Kantian Ethics' value function. Perhaps a few do, but they are surely rare exceptions. The overwhelming majority of cases of moral uncertainty fit with the bottom-up model. Very few fit with the top-down model. This constitutes another reason in favor of the bottom-up model.

III. NON-HEDGING THEORIES AND MORAL FETISHISM

So far I have shown that the problem of inter-theoretic value comparison is only threatening to hedging theories if moral decision-making is modeled as top-down. I have also suggested, though I have not yet fully developed the idea, that the fetishism problem is also only a threat to the hedging theories on the top-down model. These constitute good reasons for hedging theorists in particular to favor the bottom-up model. But I have also argued that there are independent reasons to favor the bottom-up model. The first of these is that the bottom-up model selects what is intuitively the right source for the evaluative component of moral decision-making by using the agent's direct moral evaluations of their actions, not the value function of the first order theories that they have credence in. The second independent reason to favor the bottom-up model is that it applies to a far larger subset of all of the cases that could be described as cases of moral uncertainty. These independent reasons to favor the bottom-up model put pressure on the other internalist camp, non-hedging theories, to likewise adopt the model. But I argue that the bottom-up model is not a boon for non-hedging theories in the same way it is for hedging theories. In fact, the net effect of adopting the bottom-up model is devastating for non-hedging theories.

One might think there is cause for initial optimism for the non-hedger. After all, the fetishism problem targets this view as well, and the bottom-up model shows great promise for

hedging theories in avoiding this problem. To see whether this promise materializes, let us consider the fetishism problem in more detail. Building on work by Michael Smith, Brian Weatherson argues that internalist theories require agents to have fetishistic moral motivation, because to comply with these theories, agents must care about moral compliance as such.[77] Smith argues that morally virtuous agents are not motivated by a concern to do *what they believe is right*, where this is read *de dicto*. Being motivated by compliance with morality as such would 'alienate (them) from the ends at which morality properly aims' and would make them fetishists about morality.[78] Alternatively, virtuous, non-fetishistic agents care 'non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equality, and the like'.[79]

Weatherson agrees with Smith that caring about morality as such is a kind of vicious, or least less-than-wholly-virtuous, moral motivation:

> A good person will dive into a river to rescue a drowning child . . . and won't do so because it's the right thing to do. She'll do it because there's a child who needs to be rescued, and that child is valuable.[80]

Weatherson then argues that because internalist theories direct the agent to act in light of their moral beliefs and evidence, they require an agent to act out of concern for morality as such. According to Weatherson, hedging theories require Victor to hedge, not because he cares about veal calves, but because ordering veal would be reckless, and therefore *the wrong thing to do*. Alternatively, if according to non-hedging theories Victor should not hedge, that too would be because he should be concerned with how he is behaving with respect to his moral beliefs –

[77] Weatherson (2013)
[78] Smith (1994)
[79] Smith (1994)
[80] Weatherson (2013)

instead of hedging perhaps he should be acting in compliance with the moral theory he believes to be most probable, which is common sense morality in this case. Either way, it is not the proper objects of morality that motivate Victor, according to Weatherson, but rather a concern about *getting it right*, morally.

Elsewhere I argue that these fetishism concerns are in fact a point of embarrassment for Normative Externalism because that theory cannot countenance virtuous moral motivation as being morally salient.[81] Here I want to focus on whether it succeeds against internalist theories of either the hedging or non-hedging variety on either the top-down or the bottom-up model. The first thing I argue is that Weatherson is overstating the problem when he says that internalist theories *require* one to fetishize morality. Regarding an agent who, like Victor, is morally uncertain over the moral value of livestock and is considering hedging with the vegan option, he asks:

> Why should she turn down the steak? Not because she values the interests of the cow over her dining. She does not. . . Rather, she has to care about morality as such. And that seems wrong.[82]

But internalist theories do not place any restrictions on what an agent may *care* about. They only stipulate how an agent's evaluative and probabilistic beliefs inform their moral reasons. Victor, and the agent in Weatherson's case, may care a great deal about cows, even if their credence is less than .5 that cows are morally valuable. Caring for x can vary independently of believing that x is worthy of being cared for. An expectant mother may care a great deal about the first trimester fetus gestating inside her even if her credence is less than .5 that it is in fact valuable. Of course, there is an important relationship between caring for x and believing that x is worthy

[81] 'Fetishism, Blame, and Uncertainty' (Paper 1 of this dissertation)
[82] Weatherson (2014)

of being cared for. It would be a kind of inconsistency, or at least a lack of harmony between my beliefs and desires, for me to have a great deal of moral concern for cows but have a credence of 0 that cows are valuable.[83] But Victor's credence is not 0 that veal calves matter morally. It is something like .4. So consistency, or harmony, requires that Victor's moral concern for veal calves should be less intense than if his credence was 1, but more intense than if it were 0. So merely being morally uncertain does not prevent one from having virtuous moral motivation directed at the proper objects of morality.

Moreover, there is nothing in principle about complying with an internalist theory that requires an agent to abandon that virtuous moral concern. The fact that EMVT, for example, directs agents to maximize expected moral value does not, in and of itself, mean that the theory requires fetishizing morality any more than Kantian Ethics does for directing agents to never treat others as mere means, or Act Utilitarianism does for directing agents to maximize utility. One can comply with a moral theory, whether of the first or second order, by following its directives, and also care about the proper objects of morality. EMVT does not require one to *care* about maximizing expected moral value *instead of* caring about the suffering of veal calves. It only requires one to maximize expected moral value. Of course, one *could* fetishize morality by caring only about expected moral value maximization, or at least by caring too much about it. But one could also fetishize morality by caring only about complying with the categorical imperative, or about maximizing utility. But this is not an objection to EMVT any more than it is an objection to Kantian Ethics or Act Utilitarianism.

Nevertheless, a weaker version of the fetishism problem persists. If an internalist theory directed an agent to act in such a way that was *incompatible* with the agent's virtuous moral

[83] 'Harmony' is the term used by Graham Oddie to describe the alignment of beliefs about value and desires (1994).

concern for the proper objects of morality, then that theory would be seriously suspect. To illustrate, suppose Victor really does care about the suffering of veal calves (and also about the plight of local agriculture). If a second order moral theory directed agents to flip a coin or roll dice to arbitrate their moral uncertainty, that theory would undercut Victor's virtuous moral concern. Victor's concern for the proper objects of morality simply could not be properly accounted for in his moral decision-making if he decides to order veal (or order salad) because of a coin flip. The problem is not that the coin-flipping theory of moral uncertainty *requires* Victor to *care* about coin flipping. The problem is that deciding by coin flip would disconnect whatever virtuous moral concern that Victor may have had for the proper objects of morality from his moral decision-making. To put it another way, Victor may as well not have been virtuously motivated if his decision is decided by coin flip. His moral motivation is morally pointless on the coin-flipping theory of second-order subjective moral reasons.

It is this weak version of the fetishism problem that remains daunting for hedging theories on the top-down model. Acting in light of the value functions provided by the first order moral theories one's credence ranges over, irrespective of one's direct moral evaluations of one's prospective actions is *incompatible* with caring about the proper objects of morality. Alexandra, we may presume, is virtuously motivated by the proper objects of morality – her patient and his right not to be killed to save two others. Her moral concern tracks her direct moral evaluations. But this virtuous concern for her patient is not preserved in the decision analysis that directs her to maximize utility. By maximizing utility, Alexandra might as well not be virtuously motivated, as much impact as that had on her decision-making. Maximizing utility could have preserved her virtuous motivation, were her direct concern for the proper objects of morality different from what they are. But in her case, compliance with Act Utilitarianism would

undercut her virtuous moral concern, and could only be interpreted as being motivated by moral compliance for compliance's sake. Alexandra is not *required* by hedging theories on the top-down model to *care* about moral compliance as such. But her actual virtuous moral concern is prevented from informing her subjective moral reasons on this view. This is a reason to reject hedging theories on the top-down model.

But hedging theories on the bottom-up model do not suffer from this defect. It is perfectly consistent with bottom-up hedging theories for Victor to have the following motivation: Victor cares about veal calves, but is not sure whether he should care. He also cares about supporting local agriculture. His uncertainty about whether he should care about veal calves makes him uncertain what he should do, but he judges that if veal calves are morally important, then they are very important – much more important than supporting local agriculture. At no point thus far has Victor's moral concern strayed from the proper objects of morality.

Now, Victor must decide how to act. How shall he choose in a way that preserves his concern for the proper objects of morality? I argue that he should act in such a way that his moral concern is integrated into his decision process in a principled way, reflecting both the magnitude of his moral concern and also his beliefs about the likelihood that each of his conflicting concerns accurately reflects moral reality. Plausibly, doing so will require that Victor hedge in the present circumstances, and more generally maximize his expected moral value. EMVT integrates every salient facet of Victor's moral concern – both the strength of his concern and also his subjective probability estimate over that strength.

If Victor only considered the strength of his concern, for example by using a maximin procedure which disregards his probability function, then his concern would be improperly accounted for. After all, if Victor thinks it is vanishingly unlikely that veal calves matter, then at

some point his moral concern for supporting local agriculture should outweigh his concern for veal calves – this concern counts too, but would not be taken into account if Victor used a maximin decision rule.

On the other hand, if Victor failed to integrate the magnitude of his moral concern into the decision analysis, that would likewise fail to preserve his virtuous moral concern. This is where non-hedging theories run into trouble. They are unable to countenance the magnitude of Victor's moral concern for veal calves. Doing so would make them susceptible to the problem of inter-theoretic value comparison, and collapse the difference between them and hedging theories.

Let us consider two leading non-hedging theories to see how they do with respect to our weak fetishism problem, starting with My Favorite Theory (MFT). MFT directs agents to act in accordance with the theory they consider most probable. As with the top-down understanding of EMVT, top-down MFT is incompatible with acting for virtuous moral reasons. If Victor cares about the suffering of veal calves and the plight of local agriculture, then these concerns are not adequately accounted for in the MFT decision analysis. Because Victor considers Common Sense Morality most probable, his concern for the suffering of veal calves does not enter into the decision analysis at all. But unlike for EMVT, the bottom-up model does not help MFT much at all. In a case of moral certainty, like Alexandra's, it would pick out the right moral value function (her own, not Act Utilitarianism's), and thus preserve her moral concern throughout the decision-making process. But if an agent is morally uncertain, like Victor, the agent's moral concern does not inform the decision-making process in the right way, just as in the top-down model. The magnitude, or even the existence, of Victor's moral concern for veal calves is simply not factored in, as MFT focuses exclusively on Victor's concern for local agriculture.

William MacAskill's more nuanced Borda Rule non-hedging approach does not do much better by fetishism concerns. On the top-down model, this approach directs Victor to score each of his prospective actions by their ordinal ranking according to each theory he is uncertain about. So, ordering veal might receive a 2, according to common sense morality, while ordering salad might receive a 1. Animal rights theory would score the options in reverse with a 1 for ordering veal and a 2 for ordering salad. These scores would then be weighted by his subjective probability, so that ordering veal would come out as the optimal action, because Victor's probability distribution slightly favors common sense morality. But Victor's direct moral evaluation and corresponding moral concern carries more information than the aggregation of the ordinal rankings of his prospection actions. Victor considers ordering veal to be *very* wrong if it is wrong at all. He has intense concern for the veal calves and only mild concern for supporting local agriculture. But the Borda Rule approach, on the top-down model, effaces this very important aspect of Victor's virtuous moral motivation, by focusing on the ordinal rankings generated by the two theories he is uncertain about. On the bottom-up model the Borda Rule approach also does better in cases of moral certainty. This approach prescribes that Alexandra should let the one live and the two die, corresponding to her direct (undivided) moral judgment and her virtuous moral concern. But here too, the magnitude of an agent's moral concern is discounted, and in cases of moral *uncertainty* this generates the same problems for the Borda Rule approach as in the top-down model. Victor's considerable concern for veal calves counts for no more than his much weaker concern for local agriculture.[84]

[84] It should be noted that MacAskill would likely reject using the Borda Rule approach for a case like Victor's, because Victor seems to have the requisite moral value information for maximizing expected moral value. My point here is only that if using the Borda Rule approach is fetishistic for moral uncertainty cases on the top-down model, it is also fetishistic on the bottom-up model. It is a separate question which, if any, cases are such that the Borda Rule approach is required.

The bottom-up model is not much help to non-hedging theories with respect to the weak moral fetishism problem, but the model's effect on the state of play in the literature is downright threatening for non-hedging theories due to its implications for the problem of inter-theoretic value comparison. It is not an exaggeration to say if the bottom-up model has neutralized this problem, then there is very little to recommend non-hedging theories. In fact, non-hedging theorists are quite open about the fact that the seeming intractability of the problem of inter-theoretic value comparison is a central motivation for their position, because such comparisons are not required according to their theories. It is because 'trying to weigh the importance attached by rival theories to a particular act is ultimately meaningless and fruitless,' that Edward Gracely concludes that 'the proper approach to uncertainty about the rightness of ethical theories is to determine the one most likely to be right, and to act in accord with its dictates.'[85] Johan Gustafsson and Olle Torpman write:

> Our main positive argument for MFT is that it provides consistent prescriptions over time without relying on intertheoretic comparisons of value.[86]

William MacAskill likewise begins his paper outlining and motivating his Borda Rule theory by reciting the two value comparison problems and arguing that, although these problems render hedging theories indefensible, 'decision-making in conditions of normative uncertainty and intertheoretic incomparability is not at all hopeless' thanks to his non-hedging account.[87]

Each of these non-hedging theories is best viewed as a second-best option, to be employed only if the most initially plausible theory, EMVT, is shown to be untenable. EMVT should be viewed as the default favorite among internalist theories for a couple of reasons. First,

[85] Gracely (1996)
[86] Gustafsson and Torpman (2014)
[87] MacAskill (2016)

it straightforwardly co-opts the intuitive plausibility of Expected Utility Theory, the consensus favorite approach to prudential decision-making under descriptive uncertainty, to moral decision-making under moral uncertainty. Moreover, this approach is widely favored by all parties to the moral uncertainty debate when it comes to moral decision-making under descriptive uncertainty. If Victor's uncertainty was over whether a cake was poisoned or not, no serious person would suggest that he ought to serve the cake to a friend so long as his credence just slightly favored the cake being poison-free. Because the hedging approach is so plausible in these closely related decision-making contexts, it is the presumptive right approach to moral uncertainty as well, absent some strong reason to reject it. The problem of inter-theoretic value comparison was considered to be that strong reason, but if I have succeeded, it is no longer.

The second reason to favor the hedging approach is that the non-hedging approach generates highly counter-intuitive act prescriptions in the moral uncertainty context, just as it would in descriptive uncertainty contexts. This, of course, is because it is designed to not make inter-theoretic value comparison, and so assigns no importance to moral magnitudes in the decision analysis. Consider Fred. It is the middle of the 20th century and Fred has just thought up a new, highly lucrative agricultural practice he calls 'factory farming', which will increase his company's profitability by subjecting billions of sentient animals to lives of intense suffering. Fred's credence is .51 that farm animals are morally worthless and .49 that they are as valuable as humans. Fred also has a credence of 1 that this new practice will slightly increase his own wellbeing, which he understands to give him a minor moral reason in favor of launching the new practice. We may stipulate that Fred's evidence supports his beliefs. To recap, Fred is just slightly more confident than not that launching this new agricultural practice will produce a tiny modicum of moral good and slightly less confident than not that it will unleash arguably the

greatest moral evil in human history. According to both MFT and the Borda Rule approach, Fred should launch his factory farming idea. If this were the best internalism could do, then we should despair of the prospects for an adequate internalist theory. But because the problem of inter-theoretic value comparison is no longer a threat on the bottom-up model, there is no reason to abandon the hedging approach to moral uncertainty.

IV. CONCLUSION

I have argued that the bottom-up model of moral uncertainty is superior to the top-down model in several important respects. Its adoption would have significant implications for the moral uncertainty literature. The first is that hedging theories would avoid two objections facing them in the literature, one of which is widely viewed as decisive against hedging theories, the other of which would actually become a major consideration in favor of hedging theories. The second implication is that non-hedging theories should be rejected. With the demise of the problem of intertheoretic value comparison, there is nothing to recommend these theories, and the fetishism problem remains in force against them.

BIBLIOGRAPHY

Krister Bykvist, “Evaluative Uncertainty and Consequentialist Environmental Ethics,” In *Environmental Ethics and Consequentialism*, ed. Leonard Kahn and Avram Hiller (eds.), 122-135, (2014) London: Routledge.

Jay Geyer, “Moral Uncertainty and Moral Culpability,” *Utilitas* (2018).

Edward Gracely, “On the Non-Comparability of Judgments Made by Different Ethical Theories,” *Metaphilosophy* (1996), vol. 27 no. 3.

Alexander Guerrero. “Don’t know, don’t kill: Moral ignorance, culpability, and caution,” *Philosophical Studies* (2007), vol. 136 no. 1.

Johan E. Gustafsson and Olle Torpman, “In Defence of My Favourite Theory,” *Pacific Philosophical Quarterly* (2014), vol. 95 no. 2.

Elizabeth Harman, “Does Moral Ignorance Exculpate?” *Ratio* (2011), vol. 24 no. 4.

Elizabeth Harman, “The Irrelevance of Moral Uncertainty” *Oxford Studies in Metaethics* (2015), vol. 10.

Brian Hedden, “Does MITE Make Right?: Decision-Making Under Normative Uncertainty.” *Oxford Studies in Metaethics* (2015), vol. 11.

Ted Lockhart, *Moral Uncertainty and its Consequences*. (2000), Oxford University Press, Oxford.

William MacAskill, “Moral Uncertainty as a Voting Problem,” *Mind* (2016), vol. 125 no. 500.

Graham Oddie, “Harmony, Purity, Truth,” *Mind* (1994), vol. 103 no. 412.

Graham Oddie, “Moral Uncertainty and Human Embryo Experimentation,” in *Medicine and Moral Reasoning*, ed. Fulford, K.W.M., et al, 144-161. (1994), Cambridge: Press Syndicate of the University of Cambridge.

Tom Regan, "The Case for Animal Rights," in *Advances in Animal Welfare Science* 179-189. (1987), Dordrecht: Springer.

Gideon Rosen, "Skepticism about Moral Responsibility," *Philosophical Perspectives* (2004), vol. 18 no. 1.

Jacob Ross, "Rejecting Ethical Deflationism," *Ethics* (2006), vol. 116 no. 4.

Andrew Sepielli, "What to Do When You Don't Know What to Do," *Oxford Studies in Metaethics* (2008), vol. 4.

Andrew Sepielli, "How Moral Uncertaintism Can Be Both True and Interesting," *Oxford Studies in Normative Ethics* (2017), vol. 7.

Michael Smith, *The Moral Problem* (1994), Oxford: Blackwell.

Judith Thomson, "A Defense of Abortion," *Philosophy and Public Affairs* (1971), vol. 1, no. 1.

Ralph Wedgwood, "Akrasia and Uncertainty," *Organon* (2013), vol. 20 no. 4.

Brian Weatherson, "Running Risks Morally," *Philosophical Studies* (2014), vol. 176, no. 1.