Decision-Making under Moral Uncertainty

Andrew Sepielli

University of Toronto

Abstract

Sometimes we are uncertain about matters of fundamental morality, just as we are often uncertain about ordinary factual matters. This essay considers the prospects for "moral uncertaintism" — the view that we ought to treat the first sort of uncertainty more-or-less like we treat the second. Specifically, it addresses three of the most serious worries about uncertaintism — one concerning the assignment of intermediate probabilities to moral propositions, one concerning the (im)possibility of comparing values across competing moral theories, and one concerning the possibility of higher-level normative uncertainty — i.e. not just uncertainty about what one ought to do, but uncertainty in the face of uncertainty about what one ought to do, and so on, potentially *ad infinitum*.

Introduction

Suppose that on the best scientific estimates, there is a 1-in-1000 chance that a massive asteroid is on a course to strike Earth in 20 years, wiping out humanity. Suppose further that it is possible to build a device that can re-direct the asteroid away from Earth, but that it must be built immediately for use in the next few years, in order to be effective. The cost of the device: one million dollars.

Obviously we should build such a device, even though there is a 99.9% chance we'd be wasting a million dollars to push around an asteroid that wouldn't have hit us anyway.

This doesn't settle all the questions philosophers care about, of course. Which of the many decision-rules from which this verdict follows is the correct one? Expected value maximization? Some risk-averse or risk-seeking approach? Maybe the probabilities and values are imprecise, necessitating some fancier decision theory to account for the conclusion. And what's the nature of these probabilities? Are they objective? subjective? evidential? epistemic? Are they robust facts, or mere projections of our confidence-levels? Lots for philosophers to chew on here. But again, the conclusion that we should build the device, despite the fact that it costs money and almost definitely will not be necessary, is something reasonable people can agree should guide our behaviour.

Well, just as we can be uncertain about non-moral propositions — "Will an asteroid hit the earth?" — it seems that we can be uncertain about moral ones — "What is the correct theory of punishment?"; "Is the act/omission distinction of moral significance?"; "Is meat murder?". Indeed, it's hard to make sense of moral inquiry without positing such uncertainty (Sepielli 2016). We might wonder, then, whether the kind of reasoning that applied in "asteroid" also applies in cases of moral uncertainty. Is it ever the case that, even though some moral claim may well not be true, it ought to guide our decision-making anyway because *if* it turns out to be true, then the moral cost of not acting on it is sufficiently high?

Some philosophers have argued that the answer is “yes" (Lockhart 2000; Ross 2006; Sepielli 2009; Moller 2011; Barry and Tomlin 2016; MacAskill 2016; Hicks forthcoming; Tarsney forthcoming). Maybe we should avoid killing animals for food even if there is only some chance that it is wrong. Maybe we should radically change our criminal justice system if there's any reason to suspect that retributivism is mistaken. Maybe we should donate much more of our money to charity than most of us do, since some very demanding moral theory *might* be the right one.

That's the kind of question I want to think through in this essay: However exactly we prefer to think about decision-making, even *moral* decision-making, in the face of uncertainty about the non-moral facts, should we and *can* we treat decision-making under moral uncertainty in more or less the same way? Call the view that we can and should “moral uncertaintism”. I’ll proceed by considering three worries about this view.

Worries about Probability

Very few people would object to talk of a 1-in-1000 chance of an asteroid hitting Earth, or of an 83% chance that a candidate will win the upcoming election. Many more, though, would look askance at, say, the claim that utilitarianism has a 35% chance of being true.

Some people might be worried about the *precision* of such a claim: In virtue of what is the probability of this or any other theory *exactly* 35% (or 50%, or any number)? A first pass answer is that the proponent of moral uncertaintism is by no means committed to precise probabilities. She may instead assign an imprecise probability to such a view. And there are already proposals in the literature for how to model imprecise probabilities and make decisions in light of them (Gardenfors and Sahlin 1982; Levi 1986). A more satisfying answer, though, would include an explanation of why moral propositions have the probabilities they do, whether precise or imprecise.

Such an explanation would also help us respond to those who say that the problem with assigning a 35% chance to utilitarianism is not that 35% is *precise*, but that it is *intermediate* between 0 and 1. How, they ask, can we say anything about moral claims other than that they have probability 1 if true and 0 if false? The roots of this worry are different on different interpretations of probability. If the probabilities here are so-called "modal chances", and the modality in question is metaphysical, then the problem will be that basic moral claims are *necessarily* true or false, and as such, cannot have intermediate probabilities (Mellor 2005). If they're ones to which rational or coherent agents must be able to conform their own levels of confidence, then again it's not clear that they can be intermediate, since it's arguable that any attitude regarding the correct moral theory short of certainty is *irrational.* Interpreting them as "evidential probabilities" allows for intermediate assignments only if it's right to think of basic moral claims as supported by evidence that comes in degrees (Williamson 2002; Mellor 2005). If evidence is thought of as carving up metaphysical possibility space, then there cannot, strictly speaking, by evidence of this sort for necessary propositions (Stalnaker 1984). And frequency interpretations seem unable to accommodate intermediate probabilities for basic moral claims, since these claims are *timeless*, or *atemporal.*

The easiest way out of this thicket is just to say that the probabilities here are subjective — that they're *degrees of belief, levels of confidence*, "*credences*"; that it's all in our heads. For it seems clear, as a psychological matter, that there are intermediate levels of confidence in moral claims. A person might suspect that consequentialism is true, but not be sure. I might be more confident that consequentialism is true than that a particular form of it — e.g. utilitarianism — is true, although I think utilitarianism might be true. And so on.

This answer is not *wrong* exactly, but I do think it leaves something to be desired. I don't think we should be satisfied with simply saying that there are intermediate *subjective* probabilities or levels of confidence in moral propositions. She should insist on intermediate probabilities that are mind-independent in some sense. To see why, we need to consider just what kind of theoretical and practical work a theory of decision-making under moral uncertainty is supposed to do.

Writers on the topic have identified two roles for such a theory. Some have said that it must have some connection to blame and other reactive attitudes, and their "objective correlates" — punishment and so on. This was the concern of the late-Medieval and Early Modern Catholic theologians who first formulated such theories, which they called "reflex principles" (Jonsen and Toulmin 1990). In the more contemporary literature, Alexander Guerrero (2007) rejects the view that blameless moral ignorance is necessarily exculpatory in favor of a view he calls "Don't Know, Don't Kill" — that one is blameworthy for killing something that *might* have moral status. Others have developed such theories with an eye toward their role as *guides* to action. The presupposition is that we cannot guide our behavior by norms about which we're uncertain, and so in order to take our best shot at living in accordance with our moral reasons, we need to employ some norm that in some sense takes account of the probabilities of moral claims (Sepielli 2012a).

Decision-rules that are relative to subjective probabilities are ill-suited to play either role. It would be absurd to let a Nazi off the hook for heinous acts just because he was *very confident* in the moral view upon which he based those acts (Harman 2011, 2015). What seems more relevant is how reasonable or well-grounded that confidence is. It is likewise implausible that the norms by which we fundamentally guide our behavior under uncertainty are norms that advert to that very uncertainty. For one thing, this would not fit with how we guide our actions under conditions of *certainty* or *full belief*. If I'm certain that P, I guide my behavior most fundamentally by norms like “If P, then I should do A”, not norms that advert to my having that very belief — “If *I believe that P*, then I should do A”. We don't need to look inward at our own mental states to move forward to action. It would be odd, then, to suppose that things are different in the case of uncertainty — i.e. that I fundamentally guide my behaviour by norms of the form “If *my own degree of belief* in P is thus-and-such, then I should do A”.

This last observation suggests a way forward. If the action-guiding role for agents who are certain of P is played by norms that advert to P, the same role for agents who are uncertain of P will be played by norms that advert to whatever stands in the same relation to a credence in P that P stands in to a full belief in P. So what is that relation? It’s that of *expression*. “P” *expresses* the full belief that P. By contrast, “I believe that P” *reports* a full belief that P. It’s clear that “I have a .7 degree of belief that P” reports such a degree of belief. But what expresses it? It it “There’s a .7 *objective probability* that P”? No. That’s how you express a full belief *about objective probabilities*, which we’ve already seen may not apply to fundamental moral claims.

Rather, recent writers on the topic have used the term “epistemic” for those probabilities mentioned in statements that express, rather than report, degrees of belief. (See Yalcin 2007; Swanson 2011; Moss 2013). To say that there is a .7 epistemic probability that Elizabeth Warren will win the election, or that Caligula went insane due to illness, is not to say that, in addition to Warren, the election, Caligula, illness, etc., there are these extra features of the world — epistemic probabilities. Rather, it’s simply to show (rather than to tell about) one’s own credence in those propositions, just as the straightforward claim that Warren *will* win the election would *show* one’s full belief that Warren will win.

Now, the truth of the claim that Warren will win, or that the asteroid will hit the earth, is independent of whether any person believes it, notwithstanding the fact that making such a claim expresses that belief. That’s a big difference between expressing and reporting. This is what we mean when we say that the truth or falsity of such a claim is “mind-independent”. Similarly, then, the mere fact that epistemic probability claims express (rather than report) credences does not entail that the truth of these claims *depends* on anyone’s credences. Epistemic probability claims are mind-independent, too.

This makes them better suited than claims about subjective probabilities to guide action under uncertainty, since accepting them does not involve “looking inward” at our own credences; it involves merely *having* those credences (Sepielli 2012a). It also means that the epistemic probabilities of moral propositions are relevant to blame and punishment in a way that the subjective probabilities are not (Sepielli 2017). Again, I shouldn’t be let off the hook for a heinous action just because I was very confident it was right. And finally, to return to our original question in this section, there seems to be no principled reason why intermediate probabilities of the epistemic sort could not be assigned to moral propositions. For to say that there is such-and-such an epistemic probability that some moral claim is true is not to imply anything about evidence, or frequency, or metaphysical possibility, or what the fully rational agent could think.

But to say that there is no principled reason why moral propositions might have intermediate epistemic probabilities is not yet to show affirmatively that they might have them.

To see how they might, it's helpful to turn to some examples. Consider the claim: "You ought to push the person off the bridge to stop the trolley". The probability of this claim would seem to be raised, though not to 1, by the falsity of the Doctrine of Double Effect It is raised because *one* putative ground for the wrongness of pushing the man is the DDE, and if the DDE is false, then that ground is illusory. It is not raised to 1, though, because there may be some other, genuine ground for the wrongness of pushing the man — e.g. something having to do with the causal relationships among the agent and patients, rather than with the agents' intentions (Kamm 2006).

Or consider: It seems to me that while philosophers are willing to tolerate complexity in a moral theory, they tend to assign credence in accordance with simplicity, *ceteris paribus*; a theory that cites one factor as fundamentally morally relevant is more plausible than one that cites two, and so on (Kagan 1989). Suppose that they are right in so doing. And suppose further that the well-being an action produces is at least *one* of the things that matters. It seems that these two truths raise the epistemic probability of utilitarianism, but do not confirm it beyond doubt.

Examples like this are familiar to almost everyone who has thought philosophically about ethics. Why, then, might someone nonetheless deny the existence of intermediate epistemic probabilities for moral claims?

The main worry, as far as I can see, concerns the *accessibility* of moral truths — or more specifically, their apparent *equi-accessability*. For consider that, relative to *all* of the facts, the probability of a truth is 1 and a falsehood is 0. And relative to certain collections of facts, the probability of a flipped ordinary coin landing heads is 90%, and I have a greater chance of being elected President of the United States than Warren does. To get determinate, plausible, intermediate probabilities, we need some principled way of distinguishing between the facts that may undergird an epistemic probability distribution and those that may not (Hajek 2006). The problem is that many of the ways that seem appealing in non-moral cases won't work in moral ones.

In assigning a probability to a coin landing heads or Warren winning an election, we might, say, exclude all *future* events from the supervenience base. Fundamental moral claims, however, are true or false atemporally. Or we might exclude events that are unknowable or beyond our ken — the microphysical structure of the coin being flipped, the neural structure of the brains of the voters whose behavior we aim to predict — but it's not clear that there's any analogue in the moral case. There are no moral features too small for the eye to see. Rather, it can seem that all moral claims are accessible or within our ken, especially if, as many argue, they're knowable *a priori*.

Otherwise, we might exclude claims that are *at least as hard to ascertain the truth of* as the claims whose probabilities are being assessed. The idea here is that epistemic probabilities depend on something like evidence, and evidence is something that "stands between" the thinker and the proposition the truth of which she's trying to ascertain. One ascertains such a truth *by* attending to the evidence, and so the evidence should in some sense be more accessible — an intermediary. But again, it's not obvious how this would translate to the moral realm. How would any fundamental moral claim be less accessible than any other, if indeed they're all knowable *a priori*?

To answer this question, I think we'd need to say more about how moral knowledge is gained. More specifically, we'd need to specify certain mechanisms as the ones by which we fundamentally acquire moral knowledge, and distinguish between those facts that are more readily accessible using these mechanisms, and those that are less readily accessible. While true belief about morality could in principle be arrived at in any way whatsoever, knowledge would require the employment of this mechanism. While I won't explore such a proposal here, it strikes me as plausible that some more knowledge is indeed harder to get, even though, again, all *a priori* knowledge is in *some* sense already there for the taking.

Worries about the Possibility of Intertheoretic Comparisons of Value Differences

Suppose I am uncertain whether I ought to do A or to do B. I have some credence in Theory 1, which recommends the former and some credence in Theory 2, which recommends the latter. In keeping with what we said about "asteroid", it seems that we should not simply go with the more probable theory. It seems to matter as well just how good or bad A and B are according to each of the theories — whether, e.g., there is a *big* difference between the two actions according to Theory 1, and perhaps a *smaller* difference between them on Theory 2.

But it's not obvious that these differences compare across theories (Lockhart 2000; Ross 2006; Sepielli 2009; Gustafsson and Torpman 2014; Nissan-Rozen 2015). It is not as though utilitarianism "says" "Well, the 'gap' between A and B according to me is bigger (smaller) than the gap according to deontology, which is false." This problem is similar to the problem of interpersonal comparisons of well-being that arises on preference-satisfaction conceptions thereof.

No similar problem arises in "asteroid". There the presumption is that, however morally uncertain the decision-maker is, all of the moral outlooks she finds plausible hold that the "gap" in value between "Save a million dollars & asteroid avoids the earth" and "Spend a million dollars & asteroid avoids the earth" is much smaller than the gap between the latter and "Save a million dollars & asteroid strikes the earth". There is a more or less a common scale to rank the possible outcomes, however fuzzy or imprecise. There may no such scale, though, in the case of fundamental moral uncertainty.

Several philosophers have taken heed of this problem and tried to solve it. Ted Lockhart (2000: 84) proposes something that he calls the Principle of Equity among Moral Theories (PEMT), according to which:

> ‘‘The maximum degrees of moral rightness of all possible actions in a situation according to competing moral theories should be considered equal. The minimum degrees of moral rightness of possible actions in a situation according to competing

> theories should be considered equal unless all possible actions are equally right according to one of the theories (in which case all of the actions should be considered to be maximally right according to that theory)."

But this proposal suffers from some technical difficulties (Sepielli 2012b). Most of these stem from a very general feature of Lockhart's proposal, namely that it purports to solve the problem *exogenously* — imposing a constraint on the value-assignments of the theories that is not derived from the theories themselves. It is doubtful, though, whether exogenous solutions are really solutions at all, for they do not purport to tell you how value differences according to the theories actually do, "antecedently", compare to one another. Rather, they offer a recipe for how to *impose* such comparisons, regardless of how or *whether* the differences in question really do compare.

If we opt for an exogenous solution, our reasons for action in the face of moral uncertainty will depend not only upon the probabilities of the various moral theories and how these rank actions, but also upon the particular exogenous method we've chosen. On the face of it, this seems like an unwelcome result; the last of these seems irrelevant to what we ought to do. One might reply by drawing on Lockhart's claim that the PEMT is a way of treating theories "fairly", by assigning them all equal stake in every situation. But as I've argued, this claim can't be taken at face value, since moral theories are not the kinds of things we can treat fairly or unfairly.

We should instead seek out an *endogenous* solution to the problem — one that appeals to features of the theories themselves, rather than some external constraint. I (2009) have proposed that we could compare values across theories on the basis of partial "background rankings" that they have in common. For example, suppose that I am uncertain whether to eat factory-farmed meat. On the one hand, I am certain that there is some moral benefit to doing so: it would slightly drive up wages for farm workers. On the other, I think that *maybe* mistreating animals is in the relevant sense morally equivalent to mistreating humans, and so subsidizing factory farming (FF) of cows and pigs would be tantamount to subsidizing incredibly cruel treatment of humans. While I'm uncertain about a significant question, there is a background ranking of actions in which I'm confident — one on which (A) innocuously driving up wages for unskilled workers is better than (B) buying an alternative to factory-farmed meat; on which (B) is better than (C) subsidizing cruel treatment of human beings; and one on which the "gap" in value between A and B is much smaller than the gap between B and C. We might say that I'm uncertain between two theories that "share" this background theory: Theory 1, which implies that paying for FF meat is morally equivalent to A, and Theory 2, which implies that it's tantamount to C. My thought was that this background ranking could ground a comparison between the size of the gap between eating FF meat and not on Theory 1 and the size of that gap on Theory 2. (See Figure 1.)

Figure 1 (read "~" as "is morally equivalent to")

Theory 1

A~FF B <————–—-(gap in value)——————-> C

Theory 2

A B C~FF

As Toby Ord (2008) and Brian Hedden (2015) have pointed out, this will not work. Two moral theories may agree that A is better than B, which is better than C, and even about how the A-B value-gap compares to the B-C one, without its being the case that, e.g., the A-B gap on the one theory is the same size as the A-B gap on the other. And indeed, saying otherwise leads us into contradiction. For suppose two theories rank some actions A, B, C, D and E as follows:

Figure 2

Theory 3

A B C D E

Theory 4

A B C D E

The ratio of A-B to B-C is the same on both theories, as is the ratio of C-D to D-E. But in this case, there is no way to hold A-B and B-C according to Theory 1 equal to the same value-gaps according to Theory 2 while also holding C-D and D-E equal across the two theories.

Jacob Ross (2006) has proposed an endogenous solution which is superior to mine. He proposes that we can ground intertheoretic value comparisons not merely in shared *rankings* of actions, but rather to shared *values* that underlie those rankings. In the case above, the two moral outlooks presumably "agree" not only on the A-B-C ranking but on the fundamental values of human pleasure, human pain, equality, justice, non-domination, and so on that determine this shared ranking.

Now, as it's stated, this seems to be open to the same kind of problem that plagued my proposal. Two theories may "agree" about the ratios of value differences, and may even agree on what accounts for those rankings/ratios, without its being the case that they compare intertheoretically. But I do think Ross's proposal is a step in the right direction insofar as it tries to bring *more* to bear on the problem of intertheoretic comparisons than just a set of shared rankings. It draws on features of theories that, in contexts of moral certainty, may not play any practical role over and above the rankings of actions and assignments of deontic statuses that they determine, but would be of practical importance were they capable of grounding intertheoretic comparisons.

And while, again, I don't think the substantive values to which Ross adverts will do the trick, there is another feature of moral theories that just might. We started this section with the assumption that no theory includes information about how its own value-rankings line up with those of other, by its lights false, theories. Certainly, this was no part of the utilitarianism or contractualism that we all learned about in moral philosophy class.

But I want to suggest that this assumption is actually false. Granted, nothing in the previous descriptions of Theory 1 and Theory 2 above is sufficient to fix any intertheoretic comparisons. In other words, it is consistent with everything in those descriptions that Theory 1 and Theory 2 compare to one another as represented in Figure 1, but also consistent with the descriptions that they compare as represented in this way:

Figure 3

Theory 1

A~FF B C

Theory 2

A B C

or as represented in *this* way:

Figure 4

Theory 1

A~FF B C

Theory 2

A B C~FF

Still, though, it seems that the *natural* way to commensurate the two theories or proto-theories is in the way represented in Figure 1. We imagine a university student who has never seriously thought about factory farming learning about its methods, or being exposed to the arguments against it, and coming to think that it might be on a par with the cruel treatment of human beings. Whereas her credence was once concentrated exclusively in Theory 1, now some of it shifts to Theory 2. It is odd to suppose that, along with this shift, the degree of wrongness she assigns to treating human beings cruelly (i.e. to C), on the condition that FF is tantamount to C, *diminishes* relative to the degree of wrongness she assigns to C on the condition that FF is equivalent to A instead. In other words, her new thought is not "FF and cruelty to humans might be equally bad — so I guess cruelty to humans mightn't be so bad after all!" It's "FF and cruelty to humans might be equally bad — so FF might be very bad!"

This suggests that our first-pass characterizations of Theory 1 and Theory 2 under-described the objects of at least *this* imagined student's credences. Call the pair of Theory 1 and Theory 2 as represented in Figure 1 "Pair #1", the pair as represented in Figure 3 "Pair #2"; and so on. The right thing to say about our university student is that her uncertainty is divided specifically over theories that map onto to one another as the theories in Pair #1 do. There may be no way to solve the problem of intertheoretic comparisons when the theories are Theory 1 and Theory 2 as initially described. But there is a way to solve it when the theories are those particular *versions* of Theory 1 and Theory 2 that comprise "Pair #1", which is all that matters in our imagined scenario, because *that's* what our imagined university student has in mind. The problem is soluble in this case because it it's part of the very structure of this Theory 1/Theory2 pair, as opposed to the other pairs, that certain of their value-gaps compare across the theories in this way. It's as much a part of the structure of this theory-theory *pair* as it's part of the structure of utilitarianism that you have more reason to do A than to to B if and only if A produces more utility than B does.

There are two reasons why it's easy to miss this simple solution to the problem of intertheoretic comparisons. First, it does seem like there are lots of cases where intertheoretic comparisons are impossible, or if they are possible, are so rough as to be useless in the customary sorts of moral dilemmas. We may be taking a lesson from those harder cases and cross-applying it, overzealously, to the easier cases like the one above. Second, this feature of moral theories — how their value-gaps compare to those of other theories — doesn't show up in our employment of theories in contexts of moral *certainty* or confidence. Normally, when we're assessing theories, we're considering them as items to flat-out accept or reject, in which case we care only about how they rank actions, maybe assignments of deontic statuses like "required" or "supererogatory", and maybe also their deeper explanations of why they assign the rankings and statuses that they do. We don't normally think of the kind of comparison-licensing features as parts of the theories. But that doesn't mean that they're not there, and that they can't be among the aspects of theory-structure that we have in mind when we have in mind moral theories in cases of moral uncertainty.

Worries about Higher-Order Normative Uncertainty

Just as we may be uncertain among first-order moral theories, so too might we be uncertain among theories of what to do in the face of that first-order moral uncertainty. And so there may be some pressure to posit theories yet one more level "up" — theories about what to do in the face of uncertainty about what to do in the face of first-order moral uncertainty. You can imagine how this would iterate.

I say there "*may* be" some pressure because it depends on what our grounds were in the first place for positing some norm about what to do in the face of moral uncertainty. As we saw above, these norms seem to play two roles: (1) They may be relevant to the propriety of praise, blame, reward, punishment, and so on; and (2) They may serve as guides to action. It's not so obvious to me that someone who enlists such a norm to play the first of these roles should feel any theoretical pressure to posit any rules about what to do under higher-level normative uncertainty. It's plausible that praiseworthiness and blameworthiness are a function of, among other things, whether one's action accords with rules like Guerrero's "Don't Know, Don't Kill", or with a Casustic "reflex principles" like Bartolomeo de Medina's "Probabilism" (that an action is not formally sinful so long as there's a "reasonable probability" that it is permitted) (Jonsen and Toulmin 1990: 164). The *probabilities* of these rules, unlike the probabilities of first-order moral norms, may be irrelevant to the propriety of reactive attitudes, rewards, and punishments. If that's our chief concern, then, we may have some principled basis to avoid positing higher- and higher-order norms, *ad infinitum*.

Not so if our chief concern is the guidance of action by norms. For just as we can't directly guide our conduct by first-order moral norms among which we're consciously uncertain, neither can we guide it by rules for decision-making under moral uncertainty in which we're uncertain. The search for a rule in which we can invest our full belief, and thus use as guide to action, propels us ever "upward". This is indeed *my* chief concern, and that of most contemporary writers on moral uncertainty.

The possibility of higher- and higher-order normative uncertainty gives rise to two problems. The first problem concerns normative *coherence*. Those working on this topic typically say that these decision-rules are norms of *rationality* or *subjective rightness* (Sepielli 2014). But it's not hard to imagine that they may issue different verdicts. The right thing to do in the face of moral uncertainty may differ from the right thing to do in the face of uncertainty about what to do in the face of moral uncertainty. And so on. So what's the answer to the perfectly ordinary question of what it's subjectively right, or rational, to do? Do we *aggregate* the verdicts somehow? Do we say that only the highest-order rules the agent has considered have any force — that the lower-order rules somehow lose their authority once she becomes uncertain about them?

Second, it seems that if our uncertainty in rules is in principle boundless, then it will sometimes be impossible to guide our actions by norms. In such cases, we will have to take unguided "leap of faith" in the face of our normative uncertainty, rather than accepting a norm about how to take that uncertainty into account and acting on that norm. But then these decision-rules would not be playing the role they were enlisted to play in the first place. They end up being no improvement over any first-order moral rule as it regards action-guidance. If our commitment to their truth rested on their usefulness as guides to action, we may want to go back on that commitment.

Now, these are often presented as though they're *only* problems for the moral uncertaintist: "Once you've posited a norm that's relative to the probabilities of moral claims, then you've got to keep going; there's no principled stopping point! You've crossed the Rubicon! By contrast, those who posit only norms that are relative to probabilities of non-moral propositions — e.g. the precautionary decision-rule in "asteroid" — are under no such pressure.

There is a principled stopping point — to wit, before introducing norms that advert to the probabilities of moral propositions."

But this is a mistake. First, as we just noted, if you accept moral-probability-relative norms solely because they ground blameworthiness and the like, then there may be a principled stopping point after all. Second, and more importantly, it's not at all clear why it's defensible to posit only a norm like the one in "asteroid", and then stop. If a moral uncertaintist "has to" keep going and posit higher- and higher-level norms about what to do under moral uncertainty, so "must" someone who posits a norm about what to do under non-moral uncertainty, on pain of having to take an unguided action. So as far as action-guidance is concerned, there is no principled stopping point for *anyone*. No rule is guaranteed to secure full-on acceptance, which it must if it is to fully guide action.

It seems, then, that everyone who cares about action-guidance, at least, had better be able to solve the *coherence* and *guidance* problems presented above.

In response to the *coherence* problem, I have argued that, indeed, there will be multiple verdicts about what is subjectively right or rational — with norms at each "level" relative to the probabilities of norms at the "level" below. These norms may come into conflict in the sense of recommending different verdicts, but they cannot not *appear* to come into conflict from the agent's perspective. Imagine two levels of rules: First, R1…Rn, and second, S1…Sn, which tell me what to do given the probabilities assigned to R1…Rn. Now consider: Either I am certain or I am uncertain regarding the rules in R1…Rn. Suppose first that I am *certain* that one of R1…Rn is correct. Then I can simply act on that rule, *sans* conflict from my point of view. I don't even need to consider the meta-rules in S1…Sn. And now suppose that I am *uncertain* among R1…

Rn. Then I cannot guide my conduct by any rule in R1…Rn, and must instead hope to guide it by some rule in S1…Sn. But again, there will not seem to be a conflict from my point of view. It simply cannot seem to me, from my point of view, that the R1…Rn rules guide me to do one thing, and the S1…Sn rules guide me to do something else. I will not face anything that I will regard as a practical dilemma. This is notwithstanding the fact that perhaps the *in-fact-correct* rule in R1…Rn demands that I do one thing, and the *in-fact-correct* rule in S1…Sn demands that I do something else (Sepielli 2014.)

This view gains support from an account of subjective normativity in terms of a *try* (Mason 2003; Mason 2017; Sepielli 2012a). What's right at any given level of subjective normativity is what would count as the best try at doing what one has reason, at the levels below it (including the level of first-order, or “objective” morality), to do. And while the best A-ing and the best *try* at A-ing can come apart, they cannot appear to from the agent's perspective. The latter is practically "transparent" to the former. Similar issues arise in epistemology, in connection with higher-order evidence, which includes the kind of evidence provided by peer disagreement about the force of shared, lower-order evidence (Christensen 2010; Lasonen-Aarnio 2014; Horowitz 2014; Schoenfield ms).

In response to the *guidance* problem, we should first concede that it is possible to be consciously uncertain not only about morality, but about decision-rules for moral uncertainty, decision-rules for uncertainty about those decision-rules, and so on all the way up. In some such cases, it may be that an agent cannot engage in behavior that is fully norm-guided; she will instead be consigned to take an unguided "leap of faith". But this does not show that these decision-rules have no utility as guides to action, over and above the first-order moral rules they supplement. There are two reasons for this. First, as I (2017) show, it is likely that even if I am uncertain about rules all the way up, they gradually converge in their recommendations about which *actions* to perform. And it's the latter sort of certainty that matters from the practical point of view. Second, even there is no convergence on a single action, there may be convergence on a certain disjunction of actions, at the expense of others. The higher- and higher-level norms may not converge in recommending that I do A rather than do B but may nonetheless converge in recommending that I do either of these rather than C, D, E, and so on.

Conclusion

We've surveyed three grounds for skepticism about "moral uncertaintism" — the view that we ought to respond to moral uncertainty in roughly the way we respond to uncertainty about non-normative matters. The first worry was that we may not be able to assign intermediate probabilities to moral propositions. The second worry was that there may be no way to compare values across competing moral theories. The third worry was that the possibility of higher-order normative uncertainty may threaten both moral uncertaintism's coherence and its capability to make good on its promise of providing agents with a guide to action. While there are other objections that critics have raised against the moral uncertaintist position, these three strike me as generating the *greatest* cause for concern from the *least* controversial starting-points. We've seen that the uncertaintist is not without resources to respond to these concerns, although of course these debates are far from settled.

Related Topics

Moral learning; modern moral epistemology; contemporary moral epistemology; moral expertise; moral progress and the rational resolution of disagreement

References

Barry, C., & Tomlin, P. (2016) “Moral Uncertainty and Permissibility: Evaluating Option Sets,“ *Canadian Journal of Philosophy*, 46(6), 1–26.

Christensen, D. (2010) “Higher-Order Evidence,” *Philosophy and Phenomenological Research*, 81(1), 185– 215.

Gardenfors, P. & Sahlin, N-E. (1982) “Unreliable Probabilities, Risk Taking, and Decision Making,” *Synthese*, 53(3), 361-86.

Guerrero, A. (2007) “Don’t Know, Don’t Kill: Moral Ignorance, Culpability, and Caution,” *Philosophical Studies*, 136(1), 59-97.

Gustafsson, J., & Torpman, T. (2014) “In Defence of My Favourite Theory,” *Pacific Philosophical Quarterly*, 95(2), 159–74.

Hajek, A. (2007) “The Reference Class Problem is Your Problem Too,” *Synthese*, 156(3), 563-85.

Harman, E. (2011) “Does Moral Experience Exculpate?” *Ratio* 24 (4): 443-68.

Harman, E. (2015) “The Irrelevance of Moral Uncertainty,” in R. Shafer-Landau (ed.) *Oxford Studies in Metaethics, Volume 10*, Oxford University Press.

Hedden, B. (2016) “Does MITE Make Right? On Decision-Making under Normative Uncertainty,“ in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics, Volume 11*, Oxford University Press.

Hicks, A. (forthcoming) "Moral Uncertainty and Value Comparison," in R. Shafer-Landau, *Oxford Studies in Metaethics, Volume 13*, Oxford University Press.

Horowitz, S. (2014) "Epistemic Akrasia", *Noûs*, 48 (4): 718-44.

Jonsen, A. & Toulmin S. (1990) *The Abuse of Casuistry*, University of California Press.

Kagan, S. (1989) *The Limits of Morality*, Oxford University Press.

Kamm, F.M. (2006) *Intricate Ethics*, Oxford University Press.

Lasonen-Aarnio, M. (2014) "Higher-Order Evidence and the Limit of Defeat," *Philosophy and Phenomenological Research*, 88(2), 314–45.

Levi, I. (1986) *Hard Choices*, Cambridge University Press

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

MacAskill, W. (2016) *Normative Uncertainty as a Voting Problem*, Mind

Mason, E. (2003) "Consequentialism and the 'Ought Implies Can' Principle," *American Philosophical Quarterly*, 40 (4), 319-31.

Mason, E. (2017) "Do the Right Thing: An Account of Subjective Obligation," in M. Timmons (ed.) *Oxford Studies in Normative Ethics, Volume 7*, Oxford University Press.

Mellor, D.H. (2004) *Probability: a Philosophical Introduction*, Routledge.

Moller, D. (2011) "Abortion and Moral Risk," *Philosophy*, 86(3), 425-43.

Moss, S. (2013) "Epistemology Formalized," *Philosophical Review*, 122(1), 1–43.

Nissan-Rozen, I. (2015) "Against Moral Hedging," *Economics and Philosophy*, 31(3), 1-21.

Ord, T. (2008) Personal Communication.

Ross, J. (2006) :Rejecting Ethical Deflationism," *Ethics*, 116, 742–68.

Schoenfield, M. (ms) "Two Notions of Epistemic Rationality".

Sepielli, A. (2009) “What To Do When You Don’t Know What To Do,” in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics, Volume 4*, Oxford University Press.

Sepielli, A. (2012a) “Subjective Normativity and Action Guidance,” in M. Timmons (ed.) *Oxford Studies in Normative Ethics, Volume 2*, Oxford University Press.

Sepielli, A. (2012b) “Moral Uncertainty and the Principle of Equity among Moral Theories,” *Philosophy and Phenomenological Research*, 86(3), 580–9.

Sepielli, A. (2014) “What to Do When You Don’t Know What To Do When You Don’t Know What To Do…” *Noûs*, 48(3), 521–44.

Sepielli, A. (2016) “Moral Uncertainty and Fetishistic Motivation,” *Philosophical Studies*, 173 (11), 2951–68.

Sepielli, A. (2017) “How Moral Uncertaintism Can Be Both True and Interesting,” in M. Timmons (ed.), *Oxford Studies in Normative Ethics, Volume 7*, Oxford University Press.

Stalnaker, R. (1984) *Inquiry*, Cambridge University Press.

Swanson, E. (2011) “How Not to Theorize about the Language of Subjective Uncertainty,” in A. Egan and B. Weatherson (eds.) *Epistemic Modality*, Oxford University Press.

Tarsney, C. (forthcoming) “Intertheoretic Value Comparison: a Modest Proposal,” *Journal of Moral Philosophy*.

Williamson, T. (2002) *Knowledge and Its Limits*, Oxford University Press.

Yalcin, S. (2007) “Epistemic Modals,” *Mind*, 116(464), 983–1026.

Further Readings

Greaves, H. & Ord, T. (forthcoming) “Moral Uncertainty about Population Ethics,” *Journal of Ethics and Social Philosophy*. (Applies uncertaintist reasoning to questions of population ethics.)

Enoch, D. (2014) “A Defense of Moral Deference,” *Journal of Philosophy*, 111(5), 229-58. (Defends reliance on moral experts/moral testimony partly on grounds of moral uncertainty.)

Prummer, D. (1995) *Handbook of Moral Theology*, Roman Catholic Books. (Contains a sophisticated discussion of the “reflex principles” debated by Catholic moral theologians.)

Smith, M. (2002) “Evaluation, Uncertainty, and Motivation,” *Ethical Theory and Moral Practice*, 5(3): 305-20. (Argues that non-cognitivists cannot capture the phenomenon of moral uncertainty.)

Weatherson, B. (2014) “Running Risks Morally,” *Philosophical* Studies, 167(1), 141– 63. (Argues that acting on moral-probability-relative norms requires bad motivation.)

Biography

Andrew Sepielli is Associate Professor of Philosophy at the University of Toronto. He has published papers in both normative ethics and meta-ethics, and is currently writing a book about moral objectivity.