

1

Why We Should Take Moral Uncertainty Seriously

Introduction

Our primary aim in this chapter is to introduce the topic of moral uncertainty, argue that moral uncertainty is a real and significant issue, and argue that there are non-trivial answers to the question, ‘Given that we are morally uncertain, how ought we to act in light of that uncertainty?’ We shall also consider and defend against some recent objections to the very project of trying to develop an account of decision-making under moral uncertainty: we’ll call these the *fetishism objection*; the *regress objection*; the *blameworthiness objection*; the *conscientiousness objection*; and the *disanalogy with prudence objection*.

I. Why We Should Be Morally Uncertain

The theory of how to make decisions under moral uncertainty would be fairly uninteresting if it were the case that we should always be almost certain in one particular moral view. In this section, we’ll give three arguments in favour of epistemic humility with respect to fundamental moral propositions: that, when it is contentious what the truth of a moral proposition is, we ought to be at least moderately uncertain in the truth of that moral proposition.

The Difficulty of Ethics

The first argument is simply that ethics is *hard*. As with other areas of philosophy, working out the correct moral view often involves being sensitive to subtle distinctions, being able to hold in mind many different arguments for different views, and paying attention to intuitions across many different

thought experiments. It also involves difficult questions about how to weigh different theoretical virtues, such as simplicity and elegance against intuitive plausibility. Correctly balancing all these different considerations is extremely difficult, so even when we come to a firm stance about some ethical view, we should not always expect that our reasoning is error-free. But making a mistake concerning even just one of those considerations could result in a radical change of one's view.

Moreover, we can also be morally biased in many ways, such as through vested interests, influence from one's peers, one's culture, one's religion, one's upbringing, or influence from the status quo. One can be biased towards certain views because one finds them attractive on non-epistemic grounds: perhaps a desire to quantify everything biases one against the more qualitative nuances of morality; or perhaps, to go to the other extreme, an aversion to the coldness of numbers means that one is biased against moral theories that rely on them. We can be biased as a result of evolution, or as a result of general heuristics and biases in the brain's information processing.¹

Typically, these biases are not transparent to us and, unless they've been pointed out to us, we don't know whether we suffer from them.² Moreover, these biases are generally also pernicious: even when we do know about them, we fail to adjust our beliefs adequately in response to them, and sometimes, in fact, knowledge of the bias makes the bias even worse.³ So we should assume that we're biased in many ways even if we've tried our hardest not to be. The risk of bias combined with the difficulty of ethics means that it is very easy to make an ethical mistake.

Moral Disagreement

There are some moral issues about which there is widespread agreement. For example, most of us believe that, other things being equal, it's wrong to kill an innocent adult human being. In such cases, it seems reasonable to hold that view with high confidence, despite the general difficulty of ethics. But, on

¹ Amos Tversky and Daniel Kahneman, 'Judgement under Uncertainty: Heuristics and Biases', *Science*, vol. 185, no. 4157 (27 September 1974), pp. 1124–31.

² Emily Pronin, Daniel Lin and Lee Ross, 'The Bias Blind Spot: Perceptions of Bias in Self versus Others', *Personality and Social Psychology Bulletin*, vol. 28, no. 3 (March 2002), pp. 369–81.

³ Emily Pronin, 'How We See Ourselves and How We See Others', *Science*, vol. 320, no. 5880 (30 May 2008), pp. 1177–80.

many moral issues, equally intelligent, well-read and reflective people disagree. Given this, it seems unreasonable to have very high confidence in your favoured moral view and you should think that there's at least a reasonable chance that other parties, who have thought about the issue as long as you have, have come to the correct view. This may be because these other parties have a different set of moral evidence (intuitions, life experiences, and knowledge of arguments) available to them, which, if it were available to you, would change your view. Or it may be because one of you is biased, or has made a mistake in their reasoning. Either way, given how difficult ethics is, this should reduce your confidence in your favoured moral view.

Note that the thought that one should become more uncertain in light of persistent disagreement about a difficult subject matter does not rely on any esoteric view of peer disagreement. It's a perfectly ordinary part of common-sense epistemological reasoning, and is captured by almost all views in the literature on peer disagreement.⁴

Overconfidence

The final argument is based on the fact that humans, in general, have a remarkable tendency towards overconfidence. In those cases where we can use a frequentist notion of probability, it has been repeatedly shown that we are far more confident in our views than we should be. When we give high probability estimates (above 75% likelihood of a particular event occurring or not occurring) about anything other than trivially easy matters, it's almost always true that we should have given a significantly lower estimate. This is a remarkably strong effect that is well supported by psychological evidence.⁵ Consistently, research suggests that when subjects

⁴ For example, the case of disagreement over difficult ethical issues is much more like Christensen's 'Mental Math' case, where two people quickly add up the items on a bill in their heads and each come to different totals, and therefore each person thinks it's quite likely that they might have made a mistake, than his 'Careful Checking' case, where two people have very carefully added up the items on the bill multiple times, and therefore each person thinks it's very unlikely that they have made a mistake. See David Christensen, 'Disagreement as Evidence: The Epistemology of Controversy', *Philosophy Compass*, vol. 4, no. 5 (September 2009), pp. 756–67. A view that might oppose the argument given is Thomas Kelly, 'The Epistemic Significance of Disagreement', *Oxford Studies in Epistemology*, vol. 1 (2005), pp. 167–96.

⁵ For an overview, see Sarah Lichtenstein, Baruch Fischhoff and Lawrence D. Phillips, 'Calibration of Probabilities: The State of the Art to 1980', in Daniel Kahneman, Paul Slovic, and Amos Tversky (eds), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press, 1982, pp. 306–34.

claim to be 98% certain about something, they are wrong about 30% of the time. As the degree of stated confidence gets higher, to one in a thousand probability of error, one in a million probability of error, or even '100% certainty', subjects are still wrong over 10% of the time.⁶ Moreover, this effect holds just as strongly for most experts as it does for laypersons.⁷

Because fundamental moral truths are necessarily true, we can't directly use frequency of correctness to check whether people are morally overconfident or not.⁸ But, given that we are overconfident in almost every other domain, we should expect ourselves to be overconfident in ethics, too.⁹ So it's very likely that, when we are very confident in a particular moral view, we ought to be less confident in that moral view.

These three arguments convince us that, for at least very many moral issues, we should have at least some significant degree of belief in views other than the one we favour. So moral uncertainty is a real phenomenon. The next question is whether we ought to take this uncertainty into account in our decision-making.

II. The Motivation for Taking Moral Uncertainty Seriously

There are two main motivations for believing that there are facts about how one ought to act in the face of moral uncertainty. The first is an appeal to

⁶ See Paul Slovic, Baruch Fischhoff, and Sarah Lichtenstein, 'Facts versus Fears: Understanding Perceived Risk', in Kahneman, Slovic, and Tversky (eds), *Judgment under Uncertainty*, pp. 463–90. When subjects estimated a one in a thousand probability of error, subjects were wrong 19% of the time; at one in a million odds, subjects were wrong 13% of the time. In another study, when subjects reported being '100% certain' that they were correct, they were wrong about 20% of the time. See Pauline Austin Adams and Joe K. Adams, 'Confidence in the Recognition and Reproduction of Words Difficult to Spell', *The American Journal of Psychology*, vol. 73, no. 4 (December 1960), pp. 544–52.

⁷ Sarah Lichtenstein and Baruch Fischhoff, 'Do Those Who Know More Also Know More about How Much They Know?', *Organizational Behavior and Human Performance*, vol. 20, no. 2 (December 1977), pp. 159–83.

⁸ We can still get at it through indirect means. For instance, one could use changing one's mind on an issue as a proxy for being proven wrong: if someone claims to be 99% confident in utilitarianism and then changes their mind to support egalitarianism a year later, this suggests they were overconfident.

⁹ In fact, there are grounds for supposing that the bias would be stronger in ethics. For example, overconfidence has been found to be stronger on issues that are emotionally charged: see Charles S. Taber and Milton Lodge, 'Motivated Skepticism in the Evaluation of Political Beliefs', *American Journal of Political Science*, vol. 50, no. 3 (July 2006), pp. 755–69.

our intuitions about cases involving moral dominance. Consider the following example:

Moral Dominance

Jane is at dinner, and she can either choose the foie gras, or the vegetarian risotto. Jane would find either meal equally enjoyable, so she has no prudential reason for preferring one over the other. Let's suppose that Jane finds most plausible the view that animal welfare is not of moral value so there is no moral reason for choosing one meal over another. But she also finds plausible the view that animal welfare is of moral value, according to which the risotto is the more choiceworthy option.

In this situation, choosing the risotto over the foie gras is more choiceworthy according to one hypothesis and less choiceworthy according to none. In the language of decision theory, the risotto *dominates* the foie gras. It seems very clear to us that, in some sense, Jane would act inappropriately if she were to choose the foie gras, whether or not it is morally wrong to choose the foie gras. But, if this is true, then there must be norms that take into account Jane's moral uncertainty.

A second motivation is based on the analogy with empirical uncertainty. There has been a debate concerning whether there are norms that are relative to the decision-maker's empirical credences ('subjective' norms), in addition to norms that are not relative to the decision-maker's credences ('objective' norms).¹⁰ Consider the following case developed by Frank Jackson:¹¹

Susan and the Medicine—I

Susan is a doctor, who has a sick patient, Greg. Susan is unsure whether Greg has condition A or condition C: she thinks either possibility is equally likely. And it is impossible for her to gain any evidence that will help her improve her state of knowledge any further. She has a choice of three drugs that she can give Greg: drugs A, B, and C. If she gives him drug A, and he has condition A,

¹⁰ See, for example, Frank Jackson, 'Decision-Theoretic Consequentialism and the Nearest and Dearest Objection', *Ethics*, vol. 101, no. 3 (April 1991), pp. 461–82; Michael Zimmerman, 'Is Moral Obligation Objective or Subjective?', *Utilitas*, vol. 18, no. 4 (December 2006), pp. 329–61; Peter Graham, 'In Defense of Objectivism about Moral Obligation', *Ethics*, vol. 121, no. 1 (October 2010), pp. 88–115.

¹¹ Cf. Jackson, 'Decision-Theoretic Consequentialism and the Nearest and Dearest Objection', pp. 462–3. Donald Regan gives a similar example in *Utilitarianism and Cooperation*, Oxford: Clarendon Press, 1980, pp. 264–5.

then he will be completely cured; but if she gives him drug *A*, and he has condition *C*, then he will die. If she gives him drug *C*, and he has condition *C*, then he will be completely cured; but if she gives him drug *C*, and he has condition *A*, then he will die. If she gives him drug *B*, then he will be almost completely cured, whichever condition he has, but not completely cured.

Finally, suppose that, as a matter of fact, Greg has condition *C*. So giving Greg drug *C* would completely cure him. What should Susan do? Her decision can be represented in Table 1.1.

In *some* sense, it seems that Susan ought to give Greg drug *C*: doing so is what will actually cure Greg. But given that her credence that Greg has condition *C* is only 0.5, it seems that it would be reckless for Susan to administer drug *C*. As far as she knows, in doing so she would be taking a 50% risk of Greg's death. And so it also seems that there are norms according to which the correct action for Susan is to administer drug *B*.

Similar considerations motivate the idea that there are norms that are relative to moral uncertainty. Just as one is very often uncertain about the consequences of one's actions, one is very often uncertain about which moral norms are true. Consider the following variant of the case:¹²

Susan and the Medicine—II

Susan is a doctor, who faces two sick individuals, Anne and Charlotte. Anne is a human patient, whereas Charlotte is a chimpanzee. They both suffer from the same condition and are about to die. Susan has a vial of a drug that can help. If she administers all of the drug to Anne, Anne will survive but with disability, at half the level of welfare she'd have if healthy. If Susan administers all of the drug to Charlotte, Charlotte would be completely

Table 1.1

	Greg has condition <i>A</i> —50%	Greg has condition <i>C</i> —50%
<i>A</i>	Completely cured	Dead
<i>B</i>	Almost completely cured	Almost completely cured
<i>C</i>	Dead	Completely cured

¹² This is a Jackson case under *moral* uncertainty. A similar case is given in Zimmerman, Michael. *Living with Uncertainty: The Moral Significance of Ignorance*. Cambridge: Cambridge University Press, 2008. See also, Krister Bykvist, 'Evaluative Uncertainty and Consequentialist Environmental Ethics' in Leonard Kahn and Avram Hiller (eds), *Environmental Ethics and Consequentialism*. London: Routledge, 2014, pp. 122–35 and 'How to Do Wrong Knowingly and Get Away with It' in Sliwinski Rysiek and Svensson Frans (eds), *Neither/Nor: Philosophical Papers Dedicated to Erik Carlson on the Occasion of His Fiftieth Birthday*. Uppsala: Uppsala University, 2011.

cured. If Susan splits the drug between the two, then they will both survive, but with slightly less than half the level of welfare they'd have if they were perfectly healthy. Susan is certain that the way to aggregate welfare is simply to sum it up, but is unsure about the value of the welfare of non-human animals. She thinks it is equally likely that chimpanzees' welfare has no moral value and that chimpanzees' welfare has the same moral value as human welfare. As she must act now, there is no way that she can improve her epistemic state with respect to the relative value of humans and chimpanzees.

Her three options are as follows:

- A: Give all of the drug to Anne
- B: Split the drug
- C: Give all of the drug to Charlotte

Her decision can be represented in Table 1.2, using numbers to represent how choiceworthy each option is.

While the two theories disagree strongly about the relative choiceworthiness of options *A* and *C*, they both hold that option *B* is only slightly inferior to the best option. We can represent this in terms of choiceworthiness in Table 1.3.

Finally, suppose that, according to the true moral theory, chimpanzee welfare is of the same moral value as human welfare and that therefore the morally right option is to give all of the drug to Charlotte. What should Susan do?

Table 1.2

	Anne's welfare	Charlotte's welfare
<i>A</i>	50	0
<i>B</i>	49	49
<i>C</i>	0	100

Table 1.3

	Chimpanzee welfare is of no moral value—50%	Chimpanzee welfare is of full moral value—50%
<i>A</i>	Permissible	Extremely wrong
<i>B</i>	Slightly wrong	Slightly wrong
<i>C</i>	Extremely wrong	Permissible

In the first variant of this case, under empirical uncertainty, intuitively Susan would be reckless not to administer drug *B*. Analogously, in the case above, it seems it would be morally reckless for Susan not to choose option *B*: given her credences, she would be risking severe wrongdoing by choosing either option *A* or option *C*.¹³ This motivates the idea that there are norms that take into account both one's empirical uncertainty and one's moral uncertainty.¹⁴

For these reasons, we think that there is a clear *prima facie* case in favour of the idea that there are norms that take first-order moral uncertainty into account. However, as we will see in section III, some detractors have recently expressed doubts about this idea.

III. Objections to Taking Moral Uncertainty Seriously

The worry we'll address in this section is that there is no distinctive 'ought' of decision-making under moral uncertainty. On this worry, the answer to the question of 'What ought you to do under moral uncertainty?' is just the same as the answer to the question, 'What ought you to do?', and this latter question is answered by the correct first-order moral theory, where a first-order moral theory (such as utilitarianism or Kantianism) orders options in terms of their moral choiceworthiness in a way that is not relative to moral uncertainty.¹⁵

¹³ See Bykvist, 'Evaluative Uncertainty and Consequentialist Environmental Ethics' and 'How to Do Wrong Knowingly and Get Away with It'.

¹⁴ We could strengthen this case further by considering the possibility that Susan could get more moral evidence. Suppose she knows that she will engage in sustained moral reflection in the future, and she knows that she could pay \$5 to receive a letter from her future self, which would explain her views about human versus chimpanzee welfare, and give the reasons why she holds those views. Intuitively, she ought to pay the \$5, update her moral views, and then take the best action in light of those new moral views. But, whatever the correct moral theory is, that course of action will not be the best option. On the 'human welfare only' view, giving all the drug to Anne and spending \$5 is strictly worse than simply giving the drug to Anne; on the 'all animals are equal' view, giving all the drug to Charlotte and spending \$5 is worse than simply giving all the drug to the chimpanzee. So the view on which one simply ought to do whatever is in fact morally right can't handle the intuition that, in at least some sense, it's clear that Susan ought to get more moral evidence if it's easy to do so before taking action. (Here we put aside issues of the intrinsic value of moral knowledge, or of the effects of better moral understanding on future decisions; we can assume that Susan is certain that some form of utilitarianism is correct, and that she knows that after the decision she will suffer a bout of amnesia and forget she ever made the decision or gained any moral evidence.)

¹⁵ For a statement of this worry, see Brian Weatherson, 'Review of Ted Lockhart, *Moral Uncertainty and Its Consequences*', *Mind*, vol. 111, no. 443 (July 2002), pp. 693–6 and his 'Running Risks Morally', *Philosophical Studies*, vol. 167, no. 1 (January 2014), pp. 141–63.

There are five main objections that have been raised in the literature to motivate this worry: the *fetishism objection*; the *regress objection*; the *blame-worthiness objection*; the *conscientiousness objection*; and the *disanalogy with prudence objection*. We will spend most time on the fetishism objection, in part because it is the strongest one, and in part because our answer to it will help us provide a more comprehensive view on the nature of the ‘ought’ of decision-making under moral uncertainty.¹⁶

Before we begin, we shall show why the ‘ought’ of moral uncertainty cannot be the same as the ‘ought’ of first-order moral theories. Consider, again, *Susan and the Medicine—II*. In this example, there’s one sense in which option *B* is certainly not what Susan ought to do—it’s an impermissible option on all moral views. But there’s another, different, sense, in which option *B* is what Susan ought to do. We can see that the sense of ‘ought’ must be different because, after concluding that Susan ought to choose option *B*, we do not also think that Susan should revise her credences in her moral views. Whereas if we thought that there were just one sense of ‘ought’, then her belief that she ought to choose option *B* would be inconsistent with both moral views in which she has credence.

One way of making the idea of different senses of ‘ought’ more precise is by thinking about the different senses as different *levels* of moral ought. When we face a moral problem, we are asking what we morally ought to do, *at the first level*. Standard moral theories, such as utilitarianism, Kantianism, and virtue ethics provide answers to this question. In a case of moral uncertainty, we are moving up one level and asking about what we ought to do, at the second level, when we are not sure what we ought to do at the first level. At this second level, we take into account our credence in various hypotheses about what we ought to do at the first level and what these hypotheses say about the choiceworthiness of each action. That there is such a second level moral ought seems supported by the fact that agents are morally criticizable when they, knowing all the relevant empirical facts, do what they think is very likely to be morally wrong when there is another option that is known not to pose any risk of wrong-doing. (We will give a more detailed account of this kind of moral ought in ‘The Fetishism Objection’, and levels of ought will be discussed in ‘The Regress Objection’.)

¹⁶ For an independent set of arguments along similar lines, see Andrew Sepielli, ‘Moral Uncertainty and Moralistic Motivation’, *Philosophical Studies*, vol. 173, no. 11 (November 2016), pp. 2951–68.

An alternative way to understand the ‘ought’ of moral uncertainty is in terms of *rationality*.¹⁷ Rationality, in one important sense at least, has to do with what one should do or intend, *given one’s beliefs and preferences*. This is the kind of rationality that decision theory is often seen as invoking. It can be spelled out in different ways. One is to see it as a matter of *coherence*: it is rational to do or intend what coheres with one’s beliefs and preferences.¹⁸ Another way to spell it out is to understand it as a matter of rational *processes*: it is rational to do or intend what would be the output of a rational process, which starts with one’s beliefs and preferences.¹⁹ To apply the general idea to moral uncertainty, we do not need to take a stand on which version is correct. We only need to consider agents who are morally conscientious in the following sense. They try their best to find out what is right and what is wrong. They care about doing right and refraining from doing wrong. They thus prefer doing right to doing wrong and are indifferent between different right-doings (when none of the right-doings are morally supererogatory). They also care more about serious wrong-doings than minor wrong-doings. We take this to be a precisification of the ordinary notion of conscientiousness, loosely defined as ‘governed by one’s inner sense of what is right’, or ‘conforming to the dictates of conscience’.

The idea is then to apply traditional decision-theoretical principles, according to which rational choice is some function of the agent’s preferences (utilities) and beliefs (credences). Of course, different decision-theories provide different principles (and require different kinds of utility information). But the plausible ones at least agree on cases where one option *dominates* another. Go back to the *Moral Dominance* case. Recall that Jane is considering only two moral theories: one that we may call ‘business as usual’, according to which it is permissible to eat foie gras and also permissible to eat vegetarian risotto, and another that we may call ‘vegetarianism’, according to which it is impermissible to eat foie gras but permissible to eat vegetarian risotto. The option of eating vegetarian risotto will dominate the option of eating foie gras in terms of her own preferences about right- and wrong-doings.

¹⁷ See, for instance, Sepielli, ‘What to Do When You Don’t Know What to Do When You Don’t Know What to Do...’, *Noûs*, vol. 48, no. 3 (September 2014), pp. 521–44.

¹⁸ See John Broome, *Rationality through Reasoning*, Chichester, West Sussex: Wiley-Blackwell, 2013. For a critic, see Nomy Arpaly, ‘On Acting Rationally against One’s Best Judgement’, *Ethics*, vol. 110, no. 3 (April 2000), pp. 488–513.

¹⁹ Niko Kolodny, ‘State or Process Requirements?’, *Mind*, vol. 116, no. 462 (April 2007), pp. 371–85.

No matter which moral theory is true, by eating vegetarian risotto she will ensure an outcome that she *weakly prefers* to the alternative outcome: if ‘vegetarianism’ is true, she prefers the outcome; if ‘business as usual’ is true, she is indifferent between the outcomes. The rational thing for her to do is thus to eat vegetarian risotto, given her beliefs and her preferences.

It is important to note that this decision-theoretical account of the ‘ought’ of moral uncertainty is only one of many ways of making sense of this ‘ought’. The discussion in the following chapters does not hinge on this particular account. The reason we chose to develop this account further here is that it appeals to fairly uncontroversial notions that have significance outside the debate about moral uncertainty, namely decision theoretical rationality and moral conscientiousness.

The Fetishism Objection

One might object here that we have depicted the conscientious agent as a *moral fetishist*, someone who only cares about rightness and wrongness as such rather than what makes actions right or wrong. A conscientious agent should care about helping the needy, keeping promises, and not be concerned with doing the right thing as such.²⁰

We do not think this objection is convincing, for a number of reasons. First of all, even if we concede that someone who cares about rightness and wrongness as such is a moral fetishist, it is still true that actions can be more or less rational for such agents. More generally, even immoral agents, with immoral preferences, can do what is rational, given their beliefs and immoral preferences. So, there is something these agents rationally should do when they are morally uncertain, which does not always coincide with what is in fact morally right. Hence, the fetishism objection, even if it shows

²⁰ This objection is presented in Weatherson, ‘Running Risks Morally’. The general idea that it is fetishist to care about rightness and wrongness as such is presented in Michael Smith, *The Moral Problem*, London: Wiley-Blackwell, 1994, sect. 3.5. For a critical discussion of this idea, see Nomy Arpaly, *Unprincipled Virtue: An Inquiry into Moral Agency*, New York: Oxford University Press, 2003, ch. 2, and Julia Markovits, ‘Acting for the Right Reasons’, *The Philosophical Review*, vol. 119, no. 2 (April 2010), pp. 201–42. For a response to these criticisms, see Paulina Sliwa, ‘Moral Worth and Moral Knowledge’, *Philosophy and Phenomenological Research*, vol. 93, no. 2 (September 2016), pp. 393–418.

Table 1.4

	T_1	T_2
A	Right	Right
B	Right	Wrong

that agents concerned with moral rightness are unacceptably fetishistic, does not show that moral uncertainty should never be taken seriously.

Now, it is true that the relevance of taking moral uncertainty seriously would be weakened if only morally problematic agents could take moral uncertainty seriously. But we do not think we should concede that caring about rightness and wrongness is morally problematic. To see this, we only need to consider cases where the agent has *no clue* about the non-normative descriptive features of the outcomes of the feasible actions. Suppose the choice situation is simply the following, with no defined credences for the moral hypotheses under consideration (see Table 1.4).

Here we assume that the agent has only the information depicted in this diagram. This means that the agent does not have any clue about the non-normative descriptive features of the outcomes of the actions, except that they are not exactly identical (after all, one moral theory, T_2 , deems the actions different in moral status, which precludes their outcomes being exactly similar in descriptive features). Suppose that *A* and *B* are in themselves obviously morally innocent actions, such as pressing different buttons. Under T_1 , *A* and *B* have the same choiceworthiness (and assume that it is not a case of supererogation), but T_1 does not say why both *A* and *B* are right and also equally choiceworthy. Under T_2 , *A* is right and *B* is wrong, but T_2 does not say why this is so. Even in such an informationally impoverished case, we still expect the morally conscientious agent to form preferences over the possible outcomes. More specifically, in this case, we expect a morally conscientious agent to be indifferent between right-doings and so be indifferent between (*A*, T_1) and (*B*, T_1), and prefer a right-doing to a wrong-doing and so prefer (*A*, T_2) to (*B*, T_2). The rational option, given this agent's belief and preferences, is thus to perform *A*. Performing *B* would be to risk doing wrong without any possible compensating expected gain. Now, this example is enough to show that it is simply a mistake to think that it is somehow necessarily wrong-headed, or meaningless, to ask what would be reasonable to do (in this case, rational to do) in a case of moral uncertainty.

Table 1.5

	T_1 —50%	T_2 —50%
A	Right	Major wrong
B	Minor wrong	Right

Obviously, we can go further and meaningfully ask what would be rational to do in cases where the agent lacks views about non-normative descriptive features of the relevant outcomes but has credences about various moral hypotheses. For example, take the case in Table 1.5.

A morally conscientious agent will prefer right-doings to wrong-doings and so prefer (T_1, A) to (T_1, B) and also prefer (T_2, B) to (T_2, A) . The latter preference will be stronger, since her preference for a right-doing over a *major* wrong-doing should be stronger than her preference for a right-doing over a *minor* wrong-doing. But this means that it is rational for the agent to choose B, given his beliefs and preferences, since the possible loss, from right to minor wrong, is more than compensated for by the possible gain, from major wrong to right. More generally, as a rational person he prefers bringing about the prospect $(A, 0.5, B)$ rather than $(C, 0.5, D)$, if his preference for A over C is stronger than his preference for D over B. So, just by invoking standard decision-theoretic rationality, we can easily make sense of doing what is rational in a case of moral uncertainty.

In the above examples the agent had no views about right- or wrong-making features. Will the fetishism objection come back to haunt us if we consider cases where we do have such views? We do not think so, for the objection assumes a false dichotomy: either you care about moral rightness and moral wrongness, in which case you are a moral fetishist, or you care about right-makers and wrong-makers, in which you are morally commendable. However, it is both possible and morally commendable to care about both. An agent who cares only about moral rightness seems deficient: she should also care about what she believes makes actions right, e.g. the wellbeing of affected people, promise-keeping, and truth-telling. Similarly, she does not just care about moral wrongness, but also about what she thinks makes actions wrong, e.g. the suffering of affected people, promise-breaking, and lying. And she should care about these features *for their own sake*, not just because she thinks they are right- or wrong-making. To care about these features just because they are believed to be right- or

wrong-making does indeed seem fetishist. But, equally, an agent who cares intrinsically only about these features, which she believes to make actions right or wrong, and not at all about whether her actions are right or wrong, would also be deficient as a moral agent. After all, coming to see an action as wrong should motivate a moral agent to change her intrinsic concerns so that she starts to care intrinsically about what makes actions right or wrong, according to her newly acquired moral beliefs.²¹

In the case of moral uncertainty, where the agent is uncertain about what makes actions right or wrong, it seems plausible to assume that the morally conscientious agent's intrinsic concern for the factors she believes to be right- or wrong-making will track her intrinsic concern for moral rightness and wrongness. To see what this means, consider the second example, now with added information about what the moral hypotheses say about right- and wrong-makers, which we assume are non-moral features of the options. We also assume that the agent is certain that the options have these features (see Table 1.6).

As pointed out above, a morally conscientious agent would prefer (T_1, A) to (T_1, B) , and also prefer (T_2, B) to (T_2, A) , since she prefers right-doings to wrong-doings. Furthermore, the latter preference will be stronger, since she cares more about avoiding major wrong-doings than about avoiding minor ones.

What about her intrinsic preference concerning the possible right-makers F and I , and the possible wrong-makers G and H ? Since she is morally conscientious, her intrinsic preferences over these factors will perfectly track her intrinsic preferences concerning right- and wrong-doings. Thus, she will prefer F over G and prefer I over H , and the latter preference will be stronger.²²

Table 1.6

	T_1 —50%	T_2 —50%
A	Right because of F	Major wrong because of H
B	Minor wrong because of G	Right because of I

²¹ This point is clearly stated in James Dreier, 'Dispositions and Fetishes: Externalist Models of Moral Motivation', *Philosophy and Phenomenological Research*, vol. 61, no. 3 (November 2000), pp. 619–38.

²² In the case of *radical* moral uncertainty (assuming it is possible), where the agent is uncertain about whether *one and the same* factor is either a right-maker or a wrong-maker, the intrinsic preferences concerning the right- and wrong-makers will simply match the preferences

Since it is stronger, her *overall* preference will in the end be for option *B* over option *A*.²³ Hence, there is no conflict between what is rational to prefer, given her intrinsic preferences concerning right- and wrong-makers, and what is rational to prefer, given her intrinsic preferences concerning rightness and wrongness.

What happens if the credence does not split evenly between the moral hypotheses? Again, the preferences concerning right-makers and wrong-makers will coincide with the preferences concerning rightness and wrongness. To see this, consider the case in Table 1.7.

Again, the morally conscientious person will prefer (T_1, A) to (T_1, B) , and prefer (T_2, B) to (T_2, A) , the latter preference being stronger than the former. However, since the credence speaks in favour of T_1 , to decide which option it is rational to prefer, we need to make a *trade-off* between the degree of credence and the severity of wrongness. This is just an instance of the general problem of deciding to choose between the prospect (A, p_1, B) and (C, p_2, D) , when p_1 is greater than p_2 , which speaks in favour of the former prospect, and your preference for *D* over *B* is stronger than your preference for *A* over *C*, which speaks in favour of the latter prospect. Which prospect, and corresponding option, it is rational to prefer in the end depends on how great the difference in strength is between the preferences, and how great the difference in credence is between the two moral hypotheses. We will not take a stand on how to solve this more general

Table 1.7

	T_1 —80%	T_2 —20%
A	Right because of <i>F</i>	Major wrong because of <i>H</i>
B	Minor wrong because of <i>G</i>	Right because of <i>I</i>

concerning the corresponding prospects. Suppose that in the case above, $F=H$, and $G=I$. This means that the agent is uncertain whether *F* is a right-maker or a major wrong-maker, and also uncertain about whether *G* is a minor wrong-maker or a right-maker. Her preference concerning *F* and *G* will perfectly match her preferences concerning the prospects $((T_1, A), p_1, (T_2, A))$ and $((T_1, B), p_2, (T_2, B))$. Since she prefers $((T_1, B), p_2, (T_2, B))$ to $((T_1, A), p_1, (T_2, A))$, with a certain strength, she will also intrinsically prefer *F* to *G*, with the same strength.

²³ We assume here that there are no ‘organic unities’ to take into account: if the agent prefers *A* to *B*, and *C* to *D*, he also prefers the combination of *A* and *C* (if this combination is possible) to the combination of *B* and *D* (if this combination is possible). Furthermore, we assume that the strength of the overall preference over the combinations is simply the sum of the strengths of the preferences over the combined factors.

problem here, since it requires a fuller discussion of rational preference and comparisons of moral choiceworthiness across moral theories. Here it suffices to point out that no matter how we solve the trade-off problem, it is plausible to require that, at least for fully morally conscientious agents, the overall intrinsic preferences concerning right-makers and wrong-makers should match the rational preference concerning the prospects about right- and wrong-doings.

Even if we think that the agents we have considered are morally conscientious in a perfectly normal sense, one might still question whether they are conscientious in the right way. One contestable feature of the account is that whether or not the agent's credences concern what is *in fact* right-making or wrong-making is not relevant for assessing her moral conscientiousness. But do we not want to say that a person whose intrinsic concern lines up perfectly with what is in fact right-making and wrong-making but whose moral credences are completely off track would still be morally conscientious? Some argue that Huckleberry Finn would fit this description, since he frees his friend, the slave, out of a strong feeling of compassion, even though he seems to be convinced that it is morally wrong.²⁴

Here we think it is important to distinguish between different kinds of moral appraisal. When we talk about moral conscientiousness, we have in mind a moral appraisal of the *internal* aspects of an agent, in particular, how well her preferences hang together with her moral beliefs and credences. This is not to rule out other kinds of assessments, which have more to do with external features, such as whether the agent's beliefs are true or whether her preferences line up with what is in fact morally important. We think both kinds of appraisal can live together peacefully and that there is no need to choose one over the other. Note that this kind of 'double perspective' appraisal is common in other domains. For example, a person who is deceived by an evil demon might be appraised epistemically for the way she organizes her beliefs and forms beliefs on the basis of perception, but she is still doing very poorly when it comes to truth and knowledge, two more external epistemic values.

So far, we have shown that moral uncertainty is a real issue not just for moral fetishists, but also for morally conscientious people. To this list we can also add people who are less than fully morally conscientious in the internal sense adopted above. For anyone who intrinsically cares, at least to

²⁴ This is argued in Weatherson, 'Review of Ted Lockhart, *Moral Uncertainty and Its Consequences*'.

some extent, about doing what is right and avoiding doing what is wrong will have to take moral uncertainty seriously. How much impact this will have depends on how strongly she cares about morality compared to other non-moral factors. But, since what is rational for her to do depends on her preferences, which in turn depend on her intrinsic concern for morality, what is rational for her to do will depend on her intrinsic concern for morality.²⁵ So, again, moral uncertainty needs to be taken seriously, for there is a non-trivial answer to the question of what the agent should do when faced with moral uncertainty.

It is important to add that, even if we were to concede that it is fetishistic to harbour any intrinsic attitude towards ‘thin’ moral considerations, such as rightness and wrongness, there is still room left to take moral uncertainty seriously.²⁶ Brian Weatherson, after having raised his fetishist objection, concedes that it is not fetishist to be moved by ‘thick’ moral considerations: ‘one might not do something because there is a risk that it would be cowardly, or free riding, or violate the Golden Rule, or categorical imperative.’²⁷

But this means that Weatherson has to concede that it would not be fetishistic to be guided by possible thick values when you are faced with a situation like the following: the agent is uncertain whether T_1 , a virtue theory, or T_2 , an alternative virtue theory, is correct. T_1 counts the character traits, compassion and machismo as virtues. T_2 also counts compassion as a virtue, but it counts machismo as a vice (see Table 1.8).

Table 1.8

	T_1	T_2
A	Right because A would express the virtue of compassion	Right because A would express the virtue of compassion
B	Right because B would express the virtue of machismo	Wrong because B would express the vice of machismo

²⁵ A more radical approach, which we will not pursue here, would be to drop the reference to preferences altogether and instead talk about what is rational to do, given one’s descriptive and moral beliefs and credences. Even if the agent does not care much about morality, her moral beliefs and credences can still make it rational for her to take moral uncertainty seriously. On this approach, rationality has to do with coherence between moral and descriptive beliefs and credences, on the one hand, and intentions and actions, on the other. For more on this approach, see Christian Tarsney, ‘Rationality and Moral Risk: A Defense of Moderate Hedging’, PhD thesis, University of Maryland, 2017, ch. 2.

²⁶ A similar point is also made forcefully by Sepielli, ‘Moral Uncertainty and Fetishistic Motivation’, p. 2959.

²⁷ Weatherson, ‘Running Risks Morally’, p. 159.

Here Weatherson seems willing to concede that it would be unreasonable for the agent to choose *B*, since she would risk expressing a vice. This is to concede a lot. Instead of talking about the morally conscientious agent's preference for right-doings over wrong-doings, we could say that such an agent prefers doing something that has *greater 'thick' value*: e.g. is more virtuous, satisfies rather than frustrates the Categorical Imperative, or is free-riding rather than not free-riding. The strength of her preference should match the difference in 'thick' value. With this at hand, we can again apply the standard machinery of decision theory, and determine what is rational to do for such an agent, given her preferences concerning 'thick' value.

Note that this approach will also work if one thinks that morally conscientious agents are never moved by credences in *obviously unreasonable* moral views, such as Nazi views, or fascist views.²⁸ For such agents can still be uncertain about which *reasonable* 'thick' value is weightier, e.g. whether kindness is more important than honesty. The rational choice approach can be applied to this restricted set of cases and identify what is the rational choice, given the agent's credences and preferences over reasonable 'thick' values.

Of course, one could still complain that a morally conscientious agent should only be guided by *true* 'thick' values, but, as pointed out above, that would be to completely ignore the internal aspect of agent assessment. Surely, all sides must at least agree that a morally conscientious agent can be uncertain about which *reasonable* 'thick' value is weightier and be moved by her credence about these value hypotheses.

So, we think we can conclude that moral uncertainty can be a real issue for fetishists and non-fetishists alike, and this is something that must be conceded even by people who think that only morally problematic fetishists can intrinsically care about moral rightness and wrongness. The fetishist objection to taking moral uncertainty seriously is not convincing.

One could try to resist this conclusion by arguing that, even though it is true that we can act rationally in cases of moral uncertainty without being fetishists, this is not enough to make moral uncertainty a *sufficiently* serious *moral* issue, for the following two reasons. First, since we are only talking

²⁸ Thanks to an anonymous reviewer for Oxford University Press, who pointed out that one could deny that morally conscientious agents are ever moved by moral views that are beyond the pale. Note that even if the agent tries as well as she can to find out what is right and wrong, she might still fail miserably because she has access only to very misleading evidence.

about rationality here, we seem not to be able to say that there is something *morally* problematic about risking a major moral wrong (or a major ‘thick’ disvalue).²⁹ Second, since we are talking about what is rational to do, *given* one’s preferences and beliefs, these prescriptions will only apply to agents who have preferences concerning rightness and wrongness, or ‘thick’ moral values.³⁰ How can we say that this account takes moral uncertainty sufficiently seriously, if it has nothing to say to agents who lack such preferences?

In reply, we would say that this account does in fact imply that there is something morally problematic about risking a major wrong (or a major ‘thick’ disvalue) even when you do not care about the risk. After all, if you do not care about this risk, you lack a kind of motivation that a *morally* conscientious person would have, if she shared your moral credences. But this means that you lack a motivation that is morally commendable in one respect, since being morally conscientious is one of the moral ideals. So, the account can say that there is something morally problematic about your *motivation* in this case.

The account can also say that there is something morally problematic about your *action* of taking a great moral risk, even if you do not care about it. To take such a risk is to do something that a morally conscientious and rational person would not do in your situation, if she shared your credences in moral views. Hence, it is to do something that an agent, who has a certain *moral virtue*, would not do in your situation. But to do something that an agent with a certain virtue would not do is to act in a less than fully *virtuous* way with respect to this virtue.³¹ Our account could thus be seen as providing a kind of (external) virtue assessment of actions, which is applicable even to agents who lack the preferences a virtuous agent would have.³² More exactly, your action is virtuous in this respect just in case it would be done by a morally conscientious and rational agent, who shared your credences in

²⁹ Weatherson, ‘Running Risks Morally’, p. 147.

³⁰ Michael Zimmerman, *Ignorance and Moral Obligation*, Oxford: Oxford University Press, 2014, p. 46.

³¹ This fits the standard virtue ethical formula, which can be traced back to Aristotle: an action is virtuous (in one respect) just in case a virtuous person (with a certain virtue) would do it. Another instance, famous from Kant’s writings, is the shopkeeper who always gives correct change and thus acts honestly in the sense that he does what a fully honest person would do in the circumstances. He can act honestly in this external sense and still be an egoist, who only cares about his own economic gain but realizes that giving correct change will in fact promote his own financial good.

³² This is not to deny that we can also make *internal* virtue assessments of both agents and acts. For example, an agent is kind only if she cares about other people. An act is caring only if the agent cares about others.

moral views. This account also gives us a kind of *moral ought*: what one ought to do in order to act virtuously in this respect.

It should be stressed that the relevance of the discussion to follow does not stand or fall with this virtue ethical extension of our account. Nor do we need to adopt the decision theoretical account presented above. As pointed out above, to have reason to take an interest in the discussion at all, you only need to think, with Weatherson, that there is a real issue about what to do when you are not certain which ‘thick’ values are the correct ones. Whether we should call the relevant prescriptions in cases of moral uncertainty moral (second-order), rational, virtue ethical or something else is less important. In order to signal this neutrality, we will use the term *appropriate* in the following chapters as a catch-all label for the particular normative status that is to be assigned to actions in cases of moral uncertainty.

The Regress Objection

No matter whether we see the ‘ought’ of moral uncertainty as a moral or a rational one, there seems to be a threat of an *infinite regress* of uncertainty.³³ It seems that if one can be uncertain about which first-order moral theory is correct, one can also be uncertain about how to deal with moral uncertainty itself. But it seems like this uncertainty can go even higher: one can be uncertain not only about how to deal with moral uncertainty, but also about how to deal with uncertainty about how to deal with moral uncertainty, and so on ad infinitum. We can spell this out more precisely in the following way.

Uncertainty at level 1: I am uncertain about first-order morality.

Uncertainty at level 2: I am uncertain about how to deal with uncertainty about first-order morality.

Uncertainty at level 3: I am uncertain about how to deal with uncertainty about how to deal with uncertainty about first-order morality.

³³ This kind of regress argument is stated in Weatherson, ‘Running Risks Morally’. See also Elizabeth Harman, ‘The Irrelevance of Moral Uncertainty’, *Oxford Studies in Metaethics*, vol. 10 (2015), pp. 53–79 and Sepielli, ‘How Moral Uncertainty Can be Both True and Interesting’.

Uncertainty at level 4: I am uncertain about how to deal with uncertainty about how to deal with uncertainty about how to deal with uncertainty about first-order morality.

We get an infinite regress if one's uncertainty shows up at all levels.

It is not clear that this poses a genuine threat of an *actual* infinite regress, however. First, even if we concede that for all levels i , it is possible to have uncertainty at level i , this does not show that it is possible to be in a situation in which one is uncertain at level 1, uncertain at level 2, uncertain at level 3, and so on ad infinitum. To make this inference would be to infer 'it is possible that, for all levels i , one is uncertain at level i ' from 'for all levels i , it is possible that one is uncertain at level i '. But this is to commit a simple scope fallacy: to infer 'it is possible that, for all x , Fx ' from 'for all x , it is possible that Fx '.

Second, it is not even clear that, for any level i , it is possible that one is uncertain at level i . Human agents have cognitive limitations, and therefore it is not true that they can express doubt at any level, no matter how high. There is, therefore, a natural limit on how many levels up we can go in our moral uncertainty. Indeed, just trying to figure out what uncertainty at level 4 means is tricky. In any case, it seems impossible for a human agent, in one and the same situation, to be uncertain 'all the way up' and thus have an infinite number of uncertainties.

How far should we go up in the levels of uncertainty in our theorizing? As we showed in the previous section, we cannot stay at the first level, the level of first-order morality. It is perfectly meaningful to ask about what to do in at least certain cases of moral uncertainty (minimally, the cases in which an agent is uncertain about which reasonable 'thick' value is more important). Our aim is to develop a theory for uncertainty about first-order morality.

We do not want to deny that there might be a need for a theory that can deal with *higher-order* uncertainty (higher than level 1). But our proposed theory can still fit into such a higher order theory, and this holds no matter whether the theory is 'bounded', i.e. sets a limit for the level of uncertainty that generates new prescriptions. Here is a sketchy recipe for an 'unbounded' higher order theory. Let T_1 be a first-order moral theory, such as utilitarianism, virtue ethics, or Kantianism; T_2 a theory that tells you what to do when you are uncertain about which first-order theory is correct; T_3 a theory that tells you what to do when you are uncertain

about how to deal with uncertainty about which first order theory is correct; T_4 a theory that tells you what to do when you are uncertain about how to deal with uncertainty about how to deal with uncertainty about which first-order theory is correct; and so on. One way of making this more precise is the following:

0. If you are not uncertain about T_1 , then follow T_1 .
1. If you are uncertain about T_1 and you are not uncertain about T_2 , then follow T_2 .
2. If you are uncertain about T_2 and not uncertain about T_3 , then follow T_3 .
3. If you are uncertain about T_3 and not uncertain about T_4 , then follow T_4 .

And so on. If there is no infinite regress of uncertainty, i.e. in each situation, for some level i , the agent is not uncertain about T_i , which seems to be true of all human agents, then there is always a determinate answer about what the agent should do.³⁴

Note that the theory does not require that you are *certain* that T_i is true for T_i to kick in and tell you what to do. It is enough that you fail to be certain because you do not consider T_i and thus are neither certain nor uncertain about T_i . Note also that this theory imposes *strict liability* constraints, to use Broome's apt phrase, namely the conditionals 0, 1, 2, 3, ... (etc.).³⁵ So, it cannot take into account cases of uncertainty, where the agent doubts the whole metatheory itself. But no theory can avoid imposing some strict liability constraint. Even the principle 'Deal with moral uncertainty in whatever way you see fit' is itself a principle that imposes strict liability; it tells you to deal with moral uncertainty in the way you see fit even if you don't believe that you should deal with moral uncertainty in the way you see fit.³⁶

It should be noted that the challenge of higher-order uncertainty is not just a challenge for those of us who take first-order moral uncertainty

³⁴ Obviously, for agents that are uncertain 'all the way up' there is no answer about what they should do. To take into account such unusual agents, the higher order theory needs to be revised. However, such revisions are simply beyond the scope of this book. Another bug in this theory (which we thank Christian Tarsney for alerting us to) is that it is impossible to act on a theory T_i while being less than fully certain that T_i is true. Again, revising the theory to avoid this unwelcome implication is simply beyond the scope of this book. For a more thorough discussion of different orders of normative prescriptions, see Sepielli, 'What to Do When You Don't Know What to Do When You Don't Know What to Do ...' pp. 521–44.

³⁵ Broome, *Rationality through Reasoning*, p. 75.

³⁶ This argument is inspired by Broome, *Rationality through Reasoning*, p. 101.

seriously. This is a tricky question that anyone who believes in the notion of rational choice has to confront. Even if you reject our way of dealing with moral uncertainty, you still need to know what it is rational to choose given doubts about what makes a choice rational.³⁷

To sum up, in this book, we will focus on how best to deal with uncertainty at level 1.³⁸ As pointed out above, we do not deny that there might be other levels to consider (perhaps an infinite number!), but no matter whether this is true, you need an answer about what to do when you are uncertain about which first-order moral theory is true.

The Blameworthiness Objection

Recall *Susan and the Medicine-II*. Splitting the drug between the human and the chimpanzee will guarantee that Susan performs an action that is slightly wrong. But it seems the sensible option, since the alternatives would involve risking a major wrong. If knowingly doing wrong implies blameworthiness, we have a problem, since then we then will have to say that Susan deserves blame for splitting the medicine. But we want to say that she deserves praise for not risking doing something that would have been a serious wrong.

In reply, we would maintain that the link between wrongness and blameworthiness is not this simple. There are many cases where knowingly doing wrong does not merit blame, namely, those cases in which you have a *valid excuse*.³⁹ In the Jackson cases under consideration, it seems obvious that Susan has a valid excuse; she is trying hard to figure out what is right to do but she can't tell which action is morally right. It would thus be unfair to blame her for knowingly doing wrong when she is being sensitive to her best-informed credence about degrees of wrongness and thus performing a minor wrong in order to avoid risking a major wrong. On the contrary, she

³⁷ For some discussion, see Holly Smith, 'Deciding How To Decide: Is There a Regress Problem?' in Michael Bacharach and Susan Hurley (eds), *Essays in the Foundations of Decision Theory*, Blackwell (1991), and Hanti Lin, 'On the Regress Problem of Deciding How to Decide', *Synthese* (2014).

³⁸ See MacAskill, 'Smokers, Psychos, and Decision-Theoretic Uncertainty', for a discussion of decision-making under uncertainty at level 2.

³⁹ See Cheshire Calhoun, 'Responsibility and Reproach', *Ethics*, vol. 99, no. 1 (January 1989), pp. 389–406; Gideon Rosen, 'Culpability and Ignorance', *Proceedings of the Aristotelian Society*, vol. 103, no. 1 (June 2003), pp. 61–84; Gideon Rosen, 'Skepticism about Moral Responsibility', *Philosophical Perspectives*, vol. 18, no. 1 (December 2004), pp. 295–313; and Michael Zimmerman, *Living with Uncertainty: The Moral Significance of Ignorance*, Cambridge: Cambridge University Press, 2008.

would be blameworthy if she only cared about avoiding wrong-doing and ignored morally relevant information about the differences in severity of different wrong-doings.

Some would argue that this reply is too complacent. By saying that the agent has a valid excuse because she does not know what is right, we seem also committed to the view that 'being caught in the grip of a false moral belief is exculpatory', as Elizabeth Harman puts it.⁴⁰ More precisely, we seem committed to the view that, if an agent is certain that an action is right, when in fact it is deeply wrong, she is not blameworthy for doing it. However, this view might seem wrong; surely, Hitler must be blameworthy for planning and executing the Holocaust even if he happened to be certain that it was morally right?

We agree that mere certainty cannot exculpate. But in the case of Susan, we were assuming that she did her best to decide what was right and what was wrong. Unlike Hitler, she took in all the relevant empirical information, and all the available information about moral views applicable to the situation. In the end, taking all relevant empirical and moral evidence into account, she still could not figure out which action was right. She knew which action was slightly wrong and, for each alternative action, she had 0.5 credence that it would be right and 0.5 credence that it would be deeply wrong. In order to avoid risking doing something deeply wrong, she decided to do the action she knew would be slightly wrong. This seems to be a morally commendable action and not something blameworthy. It would have been another story if she had just acted on blind faith and formed her credences about rightness and wrongness without trying her best to find out what was right and what was wrong. In this hypothetical case, we agree that it is plausible to say that Susan is not praiseworthy.

Harman seems to think that this is not enough to exculpate Susan, for she claims that:

Believing that one's wrong action is morally required involves caring inadequately about the features of one's action that make it morally wrong, because believing that an action is morally wrong on the basis of the features that make it wrong is a way of caring about those features.⁴¹

⁴⁰ Harman, 'The Irrelevance of Moral Uncertainty'. See also Alexander A. Guerrero, 'Don't Know, Don't Kill: Moral Ignorance, Culpability, and Caution', *Philosophical Studies*, vol. 136, no. 1 (October 2007), pp. 59–97 and Miranda Fricker, 'The Relativism of Blame and Williams's Relativism of Distance', *The Aristotelian Society Supplementary Volume*, vol. 84, no. 1 (June 2010), pp. 151–77.

⁴¹ Harman, 'The Irrelevance of Moral Uncertainty', p. 68.

We do not feel the force of this objection. Even if we agree that believing that an action is morally wrong on the basis of the features that make it wrong is a way of caring about those features, an agent who has done her best to identify those wrong-making features, but in the end failed, has done all we can expect from her given her unfortunate epistemic predicament. Of course, it would have been better if she could have identified the true wrong-makers. As pointed out earlier, there is always room for a more external appraisal of agents. But the fact that she could have done better in this external sense does not make her less praiseworthy.

The Conscientiousness Objection

A related objection to our account is that it violates the following constraint.

Constraint 1: It is necessarily the case that, if one acts morally conscientiously, then one does not deliberately do something that one believes to be morally wrong.

We agree that, at first sight, this constraint may sound plausible.⁴² But it is easy to conflate it with the following constraint, which we think is the true one.

Constraint 2: It is necessarily the case that, if one acts morally conscientiously, then one does not deliberately do something that one believes to be morally wrong *rather than something one believes to be morally right*.

It is easy to conflate the two constraints, since, *typically*, if an agent believes an action to be morally wrong, he also believes an alternative action to be morally right. Jackson cases such as *Susan and the Medicine—II* would be an exception. In this case, the agent knows that, in order to avoid risking a major wrong, she has to perform the action that is slightly wrong, but she does not have a belief about which action is right because she does not know which alternative action is right. We think it is clear that Susan, if she is morally conscientious, should not risk the major wrong, and thus we are willing to reject Constraint 1 and stick to Constraint 2, which is uncontroversial.

⁴² Michael Zimmerman defends this constraint under the label 'Constraint #2' (*Ignorance and Moral Obligation*, p. 33).

Disanalogy with Prudential Reasoning

One final line of objection comes from Weatherson.⁴³ The objection is that if there is an ought of moral uncertainty, then there is an ought of prudential uncertainty. But there is no ought of prudential uncertainty. Therefore, there is no ought of moral uncertainty.

We agree that if there is an ought of moral uncertainty, then there is an ought of prudential uncertainty. But we disagree with the second premise that there is no ought of prudential uncertainty. To argue for this premise, Weatherson gives the following case.

Bob and the Art Gallery

Bob has to decide whether to spend some time at an art gallery on his way home. He knows the art there will be beautiful, and he knows it will leave him cold. There isn't any cost to going, but there isn't anything else he'll gain by going either. He thinks it's unlikely that there's any prudential value in appreciating beauty, but he's not certain. As it happens, it really is true that there's no prudential value in appreciating beauty. What should Bob do?

Weatherson thinks that Bob makes no mistake in walking home. But, as is stipulated in the case, there's some chance that Bob will benefit, prudentially, from going to the art gallery, and there's no downside. This example, so the objection goes, therefore shows that there is no meaningful ought of prudential uncertainty.

We think, however, that the example is poorly chosen. Weatherson stipulates in the case that there's no cost to spending time in the art gallery. But it's difficult to imagine that to be the case: during the time that Bob would spend in the art gallery, having an experience that 'leaves him cold', could presumably have been doing something else more enjoyable instead.⁴⁴ In which case, depending on how exactly the example was specified, a plausible account of decision-making under prudential uncertainty would recommend that Bob goes home rather than to the art gallery. In order to correct for this, we could modify the case, and suppose that Bob has the choice of two routes home, *A* and *B*. Both will take him exactly the same length of time, and would require the same amount of physical effort. But route *B* passes by great works

⁴³ Weatherson, 'Running Risks Morally', pp. 148–9. For more discussion of this issue, see Christian Tarsney, 'Rationality and Moral Risk', dissertation (2017), pp. 79–83.

⁴⁴ We thank Amanda Askill for this point, and for the following example.

of architecture that Bob hasn't seen before, whereas route *A* does not. Bob knows these works are beautiful and he knows they will leave him cold. In this case, where there is some probability that viewing art has prudential value, any plausible view of decision-making under prudential uncertainty really would say it's appropriate to choose route *B* and not appropriate to choose route *A*. But that seems like the correct answer.

Other cases also suggest that there is a meaningful 'ought' of moral uncertainty in purely prudential cases. Consider the following case.⁴⁵

Charlie and the Experience Machine

Charlie is a firm believer in hedonism, but he's not certain, and gives some significant credence to the objective list theory of wellbeing. He is offered the chance to plug into the experience machine. If he plugs in, his experienced life will be much the same as it would have been anyway, but just a little bit more pleasant. However, he would be living in a fiction, and so wouldn't realize the objective goods of achievement and friendship. As it happens, hedonism is true. Is there any sense in which Charlie should not plug in?

In this case, it seems clear that there's a sense in which Charlie should not plug in. Given his uncertainty, it would be too risky for him to plug in. That is, it would be appropriate for him to refrain from plugging in, even if hedonism were true, and even if he were fairly confident, but not sure, that hedonism were true.

Or consider the following prudential Jackson case.

Pleasure or Self-Realization

Rebecca is uncertain between two theories of wellbeing, assigning them equal credence. One is hedonism, which claims that a life is good to the degree to which it is pleasant. The other is a theory which claims that a life is good to the degree to which it involves Self-Realization. She is at a pivotal life choice and has three broad options available. Option *A* is a life optimized for pleasure: she would have a decadent life with a vast amount of pleasure, but little or no Self-Realization. Option *C* is a life aimed at perfect self-realization through seclusion, study, and extreme dedication. This life would contain almost no pleasure. Option *B* is a middle path, with very

⁴⁵ We thank Amanda Askill for this case.

nearly the height of pleasure of option *A* (just one less cocktail on the beach) and very nearly the extreme Self-Realization of option *C* (just one less morning of contemplation).

It seems clear to us that given her uncertainty, the appropriate choice is for Rebecca to hedge her bets and choose option *B*.⁴⁶ While it can be a bit harder to come up with clear examples concerning prudence (due to the much greater agreement on what in practice constitutes a good life than on how to act morally), we don't see any difference in the force of the arguments in favour of there being an 'ought' of moral uncertainty whether we're considering moral uncertainty or merely prudential uncertainty.

Conclusion

In this chapter, we have seen that there is a strong case for the position that there are norms (besides first-order moral norms) that govern what we ought to do under moral uncertainty. This position is intuitive and can be made sense of by identifying these norms either with higher-level moral norms or with norms of rationality for morally conscientious agents. Moreover, we have seen that objections on the basis of fetishism, regress, blameworthiness, conscientiousness and the alleged disanalogy with prudence are unconvincing.

Having established the substance and importance of our topic, we can now start developing our positive account of what to do under moral uncertainty.

⁴⁶ The contrasting of pleasure with Self-Realization is a reference to James L. Hudson, 'Subjectivization in Ethics,' *American Philosophical Quarterly*, vol. 26, no. 3 (July 1989), p. 224, who listed these two goods as examples of values that obviously couldn't be compared. This example shows that, at least in an extreme enough case, we do have intuitions about how even these should be compared.