

6

Fanaticism and Incomparability

Introduction

In this chapter, we discuss two further problems that face accounts of decision-making under moral uncertainty, and are particularly pressing for theories that involve maximizing expected choice-worthiness.

In section I, we address the ‘fanaticism’ problem—that the expected choice-worthiness of options might be primarily determined by tiny credences in theories that posit huge amounts of value. In section II, we consider the ‘infectious incomparability’ problem—that any credence in theories with radical incomparability might render the expected choice-worthiness of almost every option undefined.

I. Fanaticism

One might worry that our account will result in *fanaticism*: that is, the expected choice-worthiness will be dominated by theories according to which most moral situations are incredibly high stakes.¹ Consider the following case.

Doug’s Lie

Doug is uncertain between two moral theories: utilitarianism, and an absolutist form of non-consequentialism. Doug has the option to tell a lie, and, in doing so, to mildly harm another person, in order to save the lives of ten people. For utilitarianism, the difference in choice-worthiness between saving ten people and saving none, all other things being equal, is 10. The difference in choice-worthiness between doing nothing and telling a lie, all other things being equal is 0.01. Absolutism agrees that it is choiceworthy to

¹ This problem was first raised by Ross, ‘Rejecting Ethical Deflationism’, p. 765.

save lives, and that it's more choiceworthy to save more lives. However, according to the absolutist, telling a lie is absolutely wrong, such that it is never permissible to tell a lie, no matter how grave the consequences. Doug is almost certain that utilitarianism is correct, but has a very small credence that the absolutist view is true.

In the above case, it seems obvious, intuitively, that it's appropriate for Doug to lie: he's almost certain both that it's the right thing to do, and that it's extremely important that he tells the lie. But, so the objection goes, this is not what MEC would recommend.

According to this objection, the most natural way to represent the absolutist theory decision-theoretically is to say that the wrong of telling a lie has infinite severity according to absolutism. If so, then, no matter how small Doug's credence is in absolutism, then the expected choice-worthiness of telling a lie is less than that of refraining from telling a lie. That is, the decision-situation looks as in Table 6.1.

If so, then, no matter how small Doug's credence is in absolutism, the expected choice-worthiness of telling a lie is less than that of refraining from telling a lie, and so refraining from lying is the appropriate option. But this seems like an absurd conclusion.

We'll consider two responses that Jacob Ross makes to this problem but then reject them and give our own response. Ross's first response is to bite the bullet, that is: 'to endorse the Pascalian conclusion, however counterintuitive it may seem at first.'² His second response is to suggest that one should not have a non-infinitesimal credence in fanatical theories:

If, therefore, one is subject to rational criticism in this case, it is not in choosing to accept [a fanatical theory] but rather in having a positive, non-infinitesimal degree of credence in a theory that is so fanatical that its

Table 6.1

	Utilitarianism—99%	Absolutism—1%
Lie	+9.99	$-\infty$
Don't lie	0	0

² Ross, 'Rejecting Ethical Deflationism', p. 766.

contribution to the expected values of one's options swamps that of all other theories.³

We cannot endorse either of these responses. Regarding the second, it is deeply implausible to claim that one should have zero credence or infinitesimal credence in any fanatical theories. We believe that absolutist theories are incorrect, but they are not so implausible as to warrant credence 0. On the standard understanding of credences,⁴ to have credence 0 in a proposition is to be certain that one could never gain any evidence that would change one's view away from credence 0. But we can clearly imagine such evidence. For example, if all our intellectual peers came to believe in absolutism after lengthy philosophical reflection, we would have reason to have positive credence in absolutism. Or if we discovered that there is a God, and His booming voice told us that absolutism is true, that would also provide evidence for absolutism. Nor, we think, does the idea of merely infinitesimal credence fare much better. First, doing so requires departing from standard Bayesianism, according to which a credence function maps onto real numbers (which does not include infinitesimals).⁵ But, second, even if we allow the possibility of rational infinitesimal credences, it seems overconfident to have such a low credence in absolutist views, despite the testimony of, for example, Kant and Anscombe, on at least some interpretations of their views. And if it's true that even some decision-makers should rationally have very small but non-infinitesimal credences in absolutist theories, then the fanaticism problem still looms large.

Regarding Ross's first response, the fanaticism problem does not merely generate grossly counterintuitive results in cases like *Doug's Lie*. Rather, it simply *breaks* MEC. In any real-life variant of *Doug's Lie*, Doug should have some non-zero credence in a view according to which it's absolutely wrong not to save those lives. In which case, the expected choice-worthiness of not lying is also negative infinity. And this will be true for any decision a real-life decision-maker faces. For any option, the decision-maker will always have some non-zero credence in a theory according to which that

³ Ross, 'Rejecting Ethical Deflationism', p. 767.

⁴ Though see Alan Hájek, 'What Conditional Probability Could Not Be', *Synthese*, vol. 137, no. 3 (December 2003), pp. 273–323 for arguments against the standard view.

⁵ For arguments against using hyperreals in our models of credences, see Kenny Easwaran, 'Regularity and Hyperreal Credences', *The Philosophical Review*, vol. 123, no. 1 (January 2014), pp. 1–41. For discussion of how invoking infinitesimals fails to help with the 'fanaticism' problem within decision theory under empirical uncertainty, see Hájek, 'Waging War on Pascal's Wager'.

option is infinitely wrong, and some non-zero credence in a theory according to which that option is infinitely right. If an option has some probability of an infinitely bad outcome, and some probability of an infinitely good outcome, then the overall expected choice-worthiness of that option will be undefined.⁶ Insofar as this is true for all options that we ever face, it means that MEC is never able to recommend one option as more appropriate than another.

A better response is simply to note that this problem arises under empirical uncertainty as well as under moral uncertainty. One should not give 0 credence to the idea that an infinitely good heaven exists, which one can enter only if one goes to church; or that it will be possible in the future through science to produce infinitely or astronomically good outcomes. This is a tricky issue within decision theory and, in our view, no wholly satisfactory solution has been provided.⁷ But it is not a problem that is unique to moral uncertainty. And we believe whatever is the best solution to the fanaticism problem under empirical uncertainty is likely to be the best solution to the fanaticism problem under moral uncertainty. This means that this issue is not a distinctive problem for moral uncertainty.

This is our primary response to the objection. However, there are, we think, two more moral uncertainty-specific things that one can say on this issue, so we briefly mention them before moving on. They both pertain to how to make comparisons of magnitudes of choice-worthiness across theories.

First, one could argue that, really, we should not understand absolutist theories as giving a quantitative measure of choice-worthiness. Instead, we should understand them as merely ordinal theories: they provide a ranking of options in terms of choice-worthiness but there is no meaning to the idea of *how much* more choiceworthy one option is than another. Absolutist theories would always rank any option that involves lying as less choiceworthy than any option that involves violating no side-constraints, but there would be no meaning to the idea that lying is 'much' more wrong than failing to save lives; there is no ratio of the difference in choice-worthiness between

⁶ For further discussion of the problems that infinite amounts of value pose for decision-theory, see Hájek, 'Waging War on Pascal's Wager'.

⁷ The standard response is to endorse prudential and moral theories whose choice-worthiness functions are bounded above and below. But this idea has severe problems of its own: making the choice-worthiness of decisions oddly dependent on facts about the past, and making bizarre recommendations when the decision-maker is close to the bound. For discussion, see Nick Beckstead, 'Recklessness, Timidity and Fanaticism', unpublished MS.

telling a lie and doing nothing and the difference in choice-worthiness between doing nothing and saving ten lives.

If so, then in accordance with the account we have defended in previous chapters, we would use the Borda Rule to aggregate our uncertainty over these theories. And if we do this, then absolutist theories would not swamp our decision-making. Our second response is that, even if one does suppose that absolutism is best represented as assigning an infinite severity of wrongness to lying, we think that the fanaticism problem is not as bad as it seems. Instead of holding that the theories agree on the choice-worthiness of saving a life, we could hold that they agree on the choice-worthiness of lying. This is still compatible with absolutism's claim that not lying is infinitely more important than saving a life, since it could treat saving a life as having a relatively infinitesimal effect on choice-worthiness—merely breaking ties in cases where the number of lies the agent told is equal. If so, then on MEC the appropriate option for Doug is to lie.⁸

Admittedly, the first way of making the intertheoretic comparison seems intuitively more plausible to us. But we're not certain that that's true. So a decision-maker like Doug should split his credence between the two different ways of making the intertheoretic comparison, giving higher credence to the one that seems more intuitively plausible. This can be spelled out more precisely, representing a theory with two kinds of choice-worthiness as a pair (c_1, c_2) in which the first element is given lexical priority, and representing the credence in the two types of normalization as credence in two types of utilitarianism: one where the choice-worthiness of promoting pleasure is treated as c_1 and one where it is treated as c_2 . If so, then Doug would have uncertainty over absolutism and two different normalizations of utilitarianism, as in Table 6.2.

Utilitarianism-1 is the normalization of utilitarianism that agrees with absolutism about the magnitude of the choice-worthiness of saving a life. Utilitarianism-2 is the normalization of utilitarianism that agrees with absolutism about the magnitude of the choice-worthiness of refraining from telling a lie. If Doug is uncertain over these two different normalizations of utilitarianism, then as long as Doug has at least one-99th as much credence in Utilitarianism-2 as he has in absolutism, MEC would recommend lying.

⁸ Christian Tarsney points out that there is a question of how this discussion interacts with the universal scale account that we defend in the previous chapter. Insofar as the idea that choice-worthiness is multidimensional is incompatible with the particular account of choice-worthiness properties that we defend in our universal scale account, we have to note that this discussion makes sense only conditional on some other metaethical view (such as that there are absolute choice-worthiness properties, but that they are multidimensional).

Table 6.2

	Utilitarianism-1	Utilitarianism-2	Absolutism-1%
Lie	(0, 9.9)	(99, 0)	(-1, 10)
Don't lie	(0, 0)	(0, 0)	(0, 0)

Taking into account uncertainty about how to normalize across theories therefore seems to get reasonably intuitive conclusions concerning what it is appropriate for one to do in real-life cases even when one has credence in what it seems initially to be a 'fanatical' moral theory.

II. Infectious Incomparability

In this book, we are largely putting aside the issue of theories that have incomplete choice-worthiness orderings. However, one might worry that in doing so we have dodged a potentially devastating problem by mere stipulation. So in this section we consider the question of whether allowing theories that posit incomparability between values, and which therefore have incomplete choice-worthiness orderings, poses an insurmountable problem for theories of decision-making under moral uncertainty.

In particular, we recast and develop further an argument taken from MacAskill, as follows.⁹ We can divide cases of incomparability into *mild* incomparability and *radical* incomparability. In cases of mild incomparability, one can sometimes (but only sometimes) make trade-offs between two different types of values. For example, perhaps you have two career paths open to you: you could be a clarinetist, or a philosopher.¹⁰ On the mild incomparability views, sometimes you can make trade-offs: if you have the option to become an outstanding clarinetist or a mediocre philosopher, then it's more choiceworthy to become the clarinetist. But, other times, such as if you have the option to become an excellent clarinetist or an excellent philosopher, there is simply no choice-worthiness relation between your

⁹ William MacAskill, 'The Infectiousness of Nihilism', *Ethics*, vol. 123, no. 3 (April 2013), pp. 508–20.

¹⁰ We take this example from Joseph Raz, *The Morality of Freedom*, Oxford: Clarendon Press, 1986, p. 332.

options: it's neither equally as choiceworthy to become the clarinetist or the philosopher, nor is one option more choiceworthy than the other.¹¹

Views that posit mild incomparability are reasonably plausible. However, we think they are unlikely to pose a grave problem for theories of decision-making under moral uncertainty. They require that we develop an account in cases of decision-making in conditions of uncertainty and incomparability; but it seems likely that one can do this in a fairly natural way. Work has already been done on this problem by Caspar Hare, for example.¹² One way of extending MEC to account for incomparability would be to claim that:

- (i) *A* is more appropriate than *B* iff *A* has greater expected choiceworthiness than *B* on all coherent completions of every moral theory in which the decision-maker has credence. (Where a choiceworthiness function CW'_i is a *coherent completion* of a moral theory T_i iff for all *A*, *B*, if *A* is at least as choiceworthy as *B* according to T_i , then $CW'_i(A) \geq CW'_i(B)$, and where the resulting choiceworthiness function has ordinal significance if the theory that is completed is ordinal, and cardinal significance if the theory that is completed is cardinal, and so on.)
- (ii) *A* is equally as appropriate as *B* iff *A* and *B* have equal expected choiceworthiness on all coherent completions of every moral theory in which the decision-maker has credence.

Let us call this the *coherent completion* account. If we took this approach, the effect will be that, given some credence in theories that posit mild incomparability, some pairs of options will be neither equally appropriate nor will one be more appropriate than the other, but most of the time (given a reasonable credence distribution) one option will be more appropriate than the other. That's a result that we can live with.

Theories that posit radical incomparability, however, are different. On these views, for some pairs of values (such as esthetic and prudential value) there are *no* trade-offs that can be made between those two types of values: any time that one option *A* increases one value by more than *B* does and *B* increases the other value by more than *A* does, then there is no positive

¹¹ Ruth Chang (ed.), *Incommensurability, Incomparability, and Practical Reason*, Cambridge, MA: Harvard University Press, 1997.

¹² See Caspar Hare, 'Take the Sugar', *Analysis*, vol. 70, no. 2 (April 2010), pp. 237–47; Riedener, 'Maximising Expected Value under Axiological Uncertainty'.

choice-worthiness relation between these two options. On many such views, there are very few positive choice-worthiness relations between options because almost all important moral decisions involve trade-offs between different types of value. This is what MacAskill calls the *infectious incomparability* problem.¹³

Views that posit radical incomparability aren't plausible, in our view. However, they are certainly epistemic possibilities, so we should assign some positive credence to them. But now suppose we try to use the coherent completion account. Because there are so few positive choice-worthiness relations, the range of possible coherent completions of a theory that posits radical incomparability is vast. This means that on our account, there will be almost no pairs of options where all the completions agree, which would be necessary for a positive appropriateness relation.

To see this, consider an example. Suppose that the decision-maker has credence in two moral views. First, she is 99.99% certain of utilitarianism. Second, she has 0.01% credence in a radical incomparabilist view on which there is no fact of the matter about how wellbeing compares between any two people; the only instances where option *A* is at least as choiceworthy than option *B* is where *A* is better for some people than *B* and worse for no one. And suppose, further, that the decision-maker is certain that there is no intertheoretic comparability between those two views. The decision-maker has the option to save one hundred lives, or to prevent the headache of a different person. On utilitarianism, let us suppose the ratio of choice-worthiness differences between *A* and *B* and *B* and *C* is 1,000,000:1. On the radical incomparabilist view *A* and *B* are incomparable in value. We can represent this as in Table 6.3.

Table 6.3

	Utilitarianism —99.99%	Radical Incomparabilism —0.01%
<i>A</i> : Save one hundred lives	1,000,000	Maximally choiceworthy
<i>B</i> : Prevent one (different person's) headache	1	Maximally choiceworthy
<i>C</i> : Do nothing	0	0

¹³ William MacAskill, 'The Infectiousness of Nihilism,' *Ethics*, vol. 123, no. 3 (April, 2013), pp. 508–20.

Clearly, the intuitively appropriate option is *A*. However, if we use the coherent completion approach, we must conclude that *A* and *B* stand in no positive appropriateness relation with each other. Any numbers can represent a coherent completion of the choice-worthiness relationship between *A* and *B*. On one coherent completion of the radical incomparabilist view, *B* is given choice-worthiness 1 trillion and *A* is given choice-worthiness 1 (which would make the expected choice-worthiness of *B* greater than that of *A*).

Our response to this problem is to point out that the argument above relies on the assumption that we can make an intertheoretic comparison between the coherent completions of the incomparabilist moral views and the other views in which the decision-maker has credence. In the example just given, we implicitly considered all possible completions of interpersonal incomparabilism *and all possible intertheoretic comparisons*. But this is not the natural way of doing things. There's no reason that we should treat the coherent completions of interpersonal incomparabilism as comparable with utilitarianism. After all, if there is rampant incomparability *within* the theory, why should we act as if there were comparability *between* it and other theories?

If, instead, we treat the coherent completions of the interpersonal incomparabilist view as incomparable with the utilitarian view, then we do not get the same infectious incomparability problem. For the purposes of working out the expected choice-worthiness of different options, we would normalize the coherent completions of infectious incomparabilism with utilitarianism at the variance of the two theories' choice-worthiness functions (which is, we argued in the last chapter, how we should in general handle theories that are incomparable with each other). If this is how we do things, then the incomparability that the theory posits is not perniciously infectious. In the above case, if we normalize the two choice-worthiness functions at their variance, there is no coherent completion of interpersonal incomparabilism such that *B* has a greater expected value than *A*.¹⁴

¹⁴ To see this, consider the coherent completion of radical incomparabilism that disagrees most strongly with utilitarianism. On this coherent completion, the choice-worthiness of *A* is epsilon greater than 0, whereas the choice-worthiness of *B* is 1. In which case, the mean of radical incomparabilism's choice-worthiness function is $\sim 1/3$ and the variance is $\sim 2/9$. Next, consider utilitarianism. Because the unit is arbitrary, we can divide the choice-worthiness values given in table 6.3 by 1,000,000 for convenience. After doing this, the mean of utilitarianism's choice-worthiness function is $\sim 1/3$ and the variance is $\sim 2/9$.

Given variance normalization, the difference in choice-worthiness between *A* and *B* on utilitarianism is therefore approximately the same as the difference in choice-worthiness between *B*

So our account would get the correct answer: that *A* is the most appropriate option.¹⁵

Conclusion

In this chapter, we discussed the fanaticism and infectious incomparability objections to accounts of decision-making under moral uncertainty. While we do not claim to have completely resolved them, we think we have showed that neither of them look like insuperable problems for our account. So let us now turn to some implications of our account of moral uncertainty.

and *A* on radical incomparabilism, on the coherent completion we're considering. Because utilitarianism has a much higher credence assigned to it, *A* will be the option with the highest expected value. Because we considered the coherent completion of radical incomparabilism that disagreed most strongly with utilitarianism, we can therefore see that under variance normalization, *A* will be the option with the highest expected choice-worthiness under all coherent completions of radical incomparabilism.

¹⁵ Christian Tarsney gave us the following objection to our account. Consider two theories: T_1 , which is classical hedonistic utilitarianism, and T_2 , which is a theory that posits both hedonic and esthetic value but holds that these two kinds of value are absolutely incomparable. Intuitively, it seems that these two views should agree on the hedonic value. So our view that we should treat these views as entirely incomparable cannot be correct.

In response, we're not wholly convinced that it is wrong to treat these views as incomparable. But, if one does find this unintuitive, there is another response one can give. In Chapter 4 we argued that the right way to make rational decisions in the face of incomparability between theories is by treating the theories as agreeing on their variance. We could broaden this account, and use it as a way of making decisions in the face of radical incomparability in general. On this view, if a theory has two value-bearers *X* and *Y* that are absolutely incomparable, then our account would, for the purposes of rational decision-making, normalize those two value-functions at their variance. If so done, then we could make the intuitive intertheoretic comparison between T_1 and T_2 above, without getting into problems with radical incomparability.