

## 2

# Maximizing Expected Choiceworthiness

### Introduction

Our primary aim in this chapter is to argue that, in conditions of interval-scale measurability and unit-comparability, one should *maximize expected choiceworthiness*. Though this position has often been suggested in the literature, and is often taken to be the ‘default’ view, it has so far received little in the way of positive argument in its favour. We start, in section I, by providing new arguments against two rival theories that have been proposed in the literature—the accounts which we call ‘My Favorite Theory’ and ‘My Favorite Option.’<sup>1</sup> Then we give a novel argument for the view that, under moral uncertainty, one should take into account both probabilities of different theories and magnitudes of choiceworthiness. Finally, we argue in favour of maximizing expected choiceworthiness (MEC).

### I. Against My Favorite Theory

One might think that, under moral uncertainty, one should simply follow the moral view that one thinks is most likely. This has been suggested as the correct principle by Edward Gracely, in one of the earliest modern papers on moral uncertainty: ‘the proper approach to uncertainty about the rightness of ethical theories is to determine the one most likely to be right, and

<sup>1</sup> We can distinguish between two versions of each of My Favorite Theory and My Favorite Option: a version which applies no matter what the informational situation of the decision-maker, and a version which applies only when theories are not comparable. We deal with the former version of these accounts here; in the next chapter we deal with the latter version. For those who are skeptical of the possibility of intertheoretic comparisons, the fact that MFT and MFO do not require intertheoretic comparisons could be considered a virtue.

to act in accord with its dictates.<sup>2</sup> Making this view more precise, we could define it as follows.

*My Favorite Theory (MFT):*  $A$  is an appropriate option iff  $A$  is a permissible option according to the theory that the decision-maker,  $S$ , has highest credence in.

This is an elegant and very simple view. But it has major problems. We'll first mention two fixable problems that need to be addressed, before moving on to a dilemma that we believe ultimately sinks the view.

The first fixable problem is that, sometimes, one will have equal highest credence in more than one moral theory. What is it appropriate to do then? Picking one theory at random seems arbitrary. So, instead, one could claim that if  $A$  is permissible according to *any* of the theories in which one has highest credence then  $A$  is appropriate. But that has odd results too. Suppose that John is 50:50 split between a pro-choice view and a radical pro-life view. According to this version of MFT, it would be appropriate for John to try to sabotage abortion clinics on Wednesday (because doing so is permissible according to the radical pro-life view) and appropriate for John to punish himself for doing so on Thursday (because doing so is permissible according to the pro-choice view). But that seems bizarre.

The second fixable problem is that the view violates the following principle, which we introduced in the previous chapter.

*Dominance:* If  $A$  is more choiceworthy than  $B$  according to some theories in which  $S$  has credence, and equally choiceworthy according to all other theories in which  $S$  has credence, then  $A$  is more appropriate than  $B$ .

MFT violates this in the case in Table 2.1.

That is, according to MFT it is equally appropriate to choose either  $A$  or  $B$ , even though  $A$  is certainly permissible, whereas  $B$  might be impermissible.

**Table 2.1**

	$T_1$ —40%	$T_2$ —60%
$A$	Permissible	Permissible
$B$	Impermissible	Permissible

<sup>2</sup> Edward J. Gracely, 'On the Noncomparability of Judgments Made by Different Ethical Theories', *Metaphilosophy*, vol. 27, no. 3 (July 1996), p. 331.

But there's no possible downside to choosing *A*, whereas there is a possible downside to choosing *B*. So it seems very plausible that it is appropriate to choose *A* and inappropriate to choose *B*.

These problems are bugs for the view, rather than fundamental objections. They can be overcome by modifying it slightly. This is what Johan Gustafsson and Olle Torpman do in a recent article.<sup>3</sup> Translating their proposal into our terminology, the version of MFT that they defend is as follows.

My Favorite Theory (Gustafsson and Torpman): An option *A* is appropriate for *S* if and only if:

1. *A* is permitted by a moral theory  $T_i$  such that
  - a.  $T_i$  is in the set  $\mathcal{T}$  of moral theories that are at least as credible as any moral theory for *S*, and
  - b. *S* has not violated  $T_i$  more recently than any other moral theory in  $\mathcal{T}$ ; and
2. There is no option *B* and no moral theory  $T_j$  such that
  - a.  $T_j$  requires *B* and not *A*, and
  - b. No moral theory that is at least as credible as  $T_j$  for *S* requires *A* and not *B*.

The first clause is designed to escape the problem of equal highest-credence theories. Clause 1(b) ensures that some bizarre courses of action are not regarded as appropriate; in the case above, if one sabotages the abortion clinic on Wednesday (following the radical pro-life view, but violating the pro-choice view), then it is not appropriate to punish oneself for doing so on Thursday (because one has violated the pro-choice view more recently than any other view). The second clause is designed to escape the problem of violating Dominance, generating a lexical version of MFT. If one's favorite theory regards all options as permissible, then one goes with the recommendation of one's second-favorite theory; if that regards all options as permissible, then one goes with the recommendation of one's third-favorite theory, and so on. This version of MFT no longer has the appeal of simplicity. But it avoids the counterintuitive results mentioned so far.

The much deeper issue with any version of MFT, however, is that it's going to run into what we'll call the *problem of theory-individuation*.

<sup>3</sup> Johan E. Gustafsson and Olle Torpman, 'In Defence of My Favourite Theory', *Pacific Philosophical Quarterly*, vol. 95, no. 2 (June 2014), pp. 159–74. Note that all of the revisions they make to their view that we discuss below are made in light of criticisms made by us in previously unpublished work or in discussion.

Consider the following case. Suppose that Sophie has credence in two different theories: a form of non-consequentialism and a form of hedonistic utilitarianism, and she's choosing between two options. *A* is the option of killing one person in order to save ten people. *B* is the option of refraining from doing so. So her decision situation is as in Table 2.2.

According to any version of MFT, *A* is the appropriate option. However, suppose that Sophie then learns of a subtle distinction between different forms of hedonistic utilitarianism. She realizes that the hedonistic theory she had credence in was actually an umbrella for two slightly different forms of hedonistic utilitarianism. So her decision situation instead looks as in Table 2.3.

In this new decision situation, according to MFT, *B* is the appropriate option. So MFT is sensitive to how exactly we choose to individuate moral theories. In order to use MFT to deliver determinate answers, we would need a canonical way in which to individuate ethical theories.

Gustafsson and Torpman respond to this with the following account of how to individuate moral theories.

Regard moral theories  $T_i$  and  $T_j$  as versions of the same moral theory if and only if you are certain that you will never face a situation where  $T_i$  and  $T_j$  yield different prescriptions.<sup>4</sup>

This avoids the arbitrariness problem, but doing so means that their view faces an even bigger problem, which is that any real-life decision-maker will

**Table 2.2**

	Non-consequentialism—40%	Utilitarianism—60%
<i>A</i>	Impermissible	Permissible
<i>B</i>	Permissible	Impermissible

**Table 2.3**

	Non-consequentialism—40%	Utilitarianism <sub>1</sub> —30%	Utilitarianism <sub>2</sub> —30%
<i>A</i>	Impermissible	Permissible	Permissible
<i>B</i>	Permissible	Impermissible	Impermissible

<sup>4</sup> Gustafsson and Torpman, 'In Defence of My Favourite Theory', p. 14.

Table 2.4

	$U$ —2%	$P_1$ —1%	$P_2$ —1%	...	$P_{98}$ —1%
A	Permissible	Impermissible	Impermissible	...	Impermissible
B	Impermissible	Permissible	Permissible	...	Permissible

have vanishingly small credence in their favorite theory. Suppose that Tracy is deciding whether to allocate resources in such a way as to provide a larger total benefit, but with an inegalitarian distribution (option *A*), or in such a way as to provide a slightly smaller total benefit, but with an egalitarian distribution (option *B*). She has some credence in utilitarianism (*U*), but is almost certain in prioritarianism (*P*). However, she's not sure exactly what shape the prioritarian weighting function should have. This uncertainty doesn't make any difference to the prioritarian recommendation in the case at hand; but it does make a small difference in some very rare cases. So her decision situation looks as in Table 2.4.

On Gustafsson and Torpman's version of MFT, the appropriate option for Tracy is *A*. But it seems intuitively obvious that it's appropriate to choose *B*, at least if we assume, as Gustafsson and Torpman do, that we cannot make choiceworthiness comparisons across theories and so we cannot appeal to the idea that there is much more at stake for the utilitarian theory than for all the prioritarian theories.

In unpublished work, Gustafsson responds to this argument. He suggests that in our argument we rely on the following principle.

*The Principle of Unconscientiousness of Almost Certain Wrongdoing:* If a morally conscientious person *P* faces a situation where options *A* and *B* are available and *P* is almost certain that *A* is wrong and almost certain that *B* is right, then *P* would not do *A*.<sup>5</sup>

Gustafsson then argues that this principle leads to choosing dominated sequences of actions.

However, our argument does not rely on this principle: indeed, this principle is inconsistent with the idea that what's appropriate is to maximize expected choiceworthiness. It is true that the account we ultimately defend can lead to intransitivity across choice-situations; we accept and defend this

<sup>5</sup> Gustafsson, 'Moral Uncertainty and the Problem of Theory Individuation' (unpublished).

implication in Chapter 4. But the issue of whether this means that one ought to choose dominated sequences of actions depends on whether a decision-maker should foresee the sequences of choices available to her and choose the *sequence* of actions that will result in the best outcome. This issue is independent from the account we defend.<sup>6</sup>

The true solution to the problem of theory individuation might seem obvious. Rather than focus on what *theory* the decision-maker has most credence in, we should instead think about what *option* is most likely to be right, in a given decision situation. That is, we should endorse something like the following.

*My Favorite Option (MFO):* *A* is an appropriate option for *S* iff *S* thinks that *A* is the option, or one of the options, that is most likely to be permissible.<sup>7</sup>

MFO isn't sensitive to how we individuate theories. And it would get the right answer in the prioritarianism and utilitarianism case above. So it looks much more plausible than MFT. But it still has a serious problem (which MFT also suffers from): it doesn't allow us to make trade-offs between the degree of credence that one has in different moral views and the degree of choiceworthiness that those views assign to different options. We'll turn to this next.

## II. In Favour of Trade-offs

We can construct examples to support the view that the correct theory of decision-making under moral uncertainty should consider trade-offs. First, suppose that your credence is split between two theories, with the second theory being just slightly more plausible. MFT and MFO both claim that you should do whatever this second theory recommends because it has the highest chance of being right. Suppose, however, that the theories disagree

<sup>6</sup> A further objection to Gustafsson and Torpman's version of *My Favorite Theory* is that the account loses the underlying motivation for thinking that there's an 'ought' that's relative to moral uncertainty in the first place. MFT, on their account, is not action-guiding. Nor can they draw support from the analogy with decision-making under empirical uncertainty. Given this, it's hard to see why we should endorse their view over the hard externalist position of Weatherston and Harman.

<sup>7</sup> Lockhart suggests this view, though ultimately rejects it (*Moral Uncertainty and Its Consequences*, p. 26).

not only on the right act but also on the magnitude of what is at stake. The slightly more plausible theory says it is a minor issue, while the less plausible one says that it is a matter of grave importance. We can represent this as in Table 2.5.

For vividness, suppose that the decision-maker is unsure about the acts/omissions distinction:  $T_1$  is the view according to which there is no morally relevant distinction between acts and omissions;  $T_2$  is the view according to which there is an important morally relevant distinction between acts and omissions. Let option  $B$  involve seriously harming many people in order to prevent a slightly greater harm to another group, while option  $A$  is keeping the status quo. Even if one is leaning slightly towards  $T_1$ , it seems morally reckless to choose  $B$  when  $A$  is almost as good on  $T_2$ 's terms and much better on  $T_1$ 's terms. Just as we can 'hedge our bets' in situations of descriptive uncertainty, so it seems that  $B$  would morally hedge our bets, allowing a small increase in the chance of acting wrongly in exchange for a greatly reduced degree of potential wrongdoing.

For a second example, consider again *Susan and the Medicine—II* (see Table 2.6).

According to MFT and MFO, both  $A$  and  $C$  are appropriate options, while  $B$  is inappropriate. But that seems wrong.  $B$  seems like the appropriate option, because, in choosing either  $A$  or  $C$ , Susan is risking grave wrongdoing.  $B$  seems like the best hedge between the two theories in which she has credence. But if so, then any view on which the appropriate option is always the maximally choiceworthy option according to some theory in which one has credence must be false. This includes MFT, MFO, and their variants.

**Table 2.5**

	$T_1$ —51%	$T_2$ —49%
$A$	Permissible	Gravely wrong
$B$	Slightly wrong	Permissible

**Table 2.6**

	Chimpanzee welfare is of no moral value—50%	Chimpanzee welfare is of significant moral value—50%
$A$	Permissible	Extremely wrong
$B$	Slightly wrong	Slightly wrong
$C$	Extremely wrong	Permissible

One might object that making trade-offs requires the possibility of intertheoretic choiceworthiness comparisons and argue that, since such comparisons are impossible, the above examples are spurious.<sup>8</sup> Our response is discussed at far greater length in Chapters 3–5: Chapter 5 argues that such comparisons are often meaningful; Chapter 4 argues that, even when they are not meaningful, we still have a principled method of placing those different moral theories on the same scale; and Chapter 3 argues that, even when the moral theories themselves provide a merely ordinal measure of choiceworthiness (and there are not meaningful ratios of choiceworthiness differences *within* a theory), we should still want to make trade-offs and MFT and MFO should be rejected. In the meantime, we will proceed on the assumption that such comparisons are meaningful.

An alternative response is suggested by Gustafsson.<sup>9</sup> Drawing on a suggestion from Tarsney,<sup>10</sup> he suggests a more coarse-grained form of My Favorite Theory: that, rather than acting in accordance with the individual moral theory in which one has the highest credence, one should instead act in accordance with the class of mutually comparable theories in which one has highest credence, and maximize expected choiceworthiness with respect to that class. Gustafsson suggests this is still a form of My Favorite Theory insofar as it is treating intertheoretically comparable theories as different specifications of the same theory.

In the next two chapters, we will argue in favour of an alternative account of what to do in varying informational conditions. For now we'll note that Gustafsson's suggestion still suffers from a grave problem for MFT that we noted earlier. Consider the utilitarianism vs prioritarianism case given above, and assume that none of the theories are comparable with each other. Coarse-grained MFT would recommend acting in accordance with utilitarianism: that is, it recommends acting in accordance with one's favorite theory even when one has vanishingly small credence in that theory, and even when all other theories oppose the recommendation of one's favorite theory.

Finally, as a side point, we note that *Susan and the Medicine—II* shows that one understanding of the central question for decision-making given moral uncertainty that has been presented in the literature by Jacob Ross, and which might lead one to find MFT or MFO attractive, is wrong. Ross seems to suggest that the central question is 'What ethical theories are

<sup>8</sup> See Gustafsson and Torpman, 'In Defence of My Favourite Theory', and Gustafsson, 'Moral Uncertainty and the Problem of Theory Individuation'.

<sup>9</sup> Gustafsson, 'Moral Uncertainty and the Problem of Theory Individuation'.

<sup>10</sup> Tarsney, 'Rationality and Moral Risk', pp. 215–19.



worthy of acceptance and what ethical theories should be rejected?', where acceptance is defined as follows.<sup>11</sup>

to accept a theory is to aim to choose whatever option this theory would recommend, or in other words, to aim to choose the option that one would regard as best on the assumption that this theory is true. For example, to accept utilitarianism is to aim to act in such a way as to produce as much total welfare as possible, to accept Kantianism is to aim to act only on maxims that one could will as universal laws, and to accept the Mosaic Code is to aim to perform only actions that conform to its Ten Commandments.

The above case shows that this cannot be the right way of thinking about things. Option *B* is wrong, according to all theories in which Susan has credence: she is certain that it's wrong. The central question is therefore not about which first-order moral theory to accept: indeed, in cases like Susan's there is *no* moral theory that she should accept. Instead, it's about which *option* it is appropriate to choose.<sup>12</sup>

### III. In Favour of Treating Moral and Empirical Uncertainty Analogously

In the previous section, we discussed an argument in favour of the view that appropriateness involves trade-offs between levels of credence in different theories and the degree of choiceworthiness that those theories assign to options. But this still leaves open exactly what account of decision-making under moral uncertainty is correct. In this section, we argue that, when choiceworthiness differences are comparable across theories, we should handle moral uncertainty in just the same way that we should handle empirical uncertainty. Expected utility theory is the standard account of how to handle empirical uncertainty probabilities.<sup>13</sup> So maximizing expected

<sup>11</sup> Jacob Ross, 'Rejecting Ethical Deflationism', *Ethics*, vol. 116, no. 4 (July 2006), p. 743.

<sup>12</sup> One could say that, in Susan's case, she should accept a theory that represents a hedge between the two theories in which she has credence (cf. Ross, 'Rejecting Ethical Deflationism', pp. 743–4). But why should she accept a theory that she knows to be false? This seems to be an unintuitive way of describing the situation, for no additional benefit.

<sup>13</sup> At least, expected utility theory is the correct account of how to handle empirical uncertainty when we have well-defined probabilities over states of nature. As we noted in the introduction, in this book we're assuming that we have well-defined credences over moral theories. If we had,

choiceworthiness should be the standard account of how to handle moral uncertainty.<sup>14</sup> This provides a further argument against MFT and MFO, which break from this standard approach.<sup>15</sup>

We can thus define the following rival to MFT and MFO:

*Maximize Expected Choiceworthiness (MEC):* When we can determine the expected choiceworthiness of different options, *A* is an appropriate option iff *A* has the maximal expected choiceworthiness.

The argument for treating empirical and moral uncertainty analogously begins by considering that there are very many ways of distinguishing between proposition-types: we can divide propositions into the a priori and a posteriori, the necessary and contingent, or those that pertain to biology and those that do not.<sup>16</sup> These could all feature into uncertainty over states of nature. Yet, intuitively, in all these cases the nature of the propositions over which one is uncertain does not affect which normative theory we should use. So it would seem arbitrary to think that *only* in the case of normative propositions does the nature of the propositions believed affect which decision-theory is relevant. So it seems like the default view is that moral and empirical uncertainty should be treated in the same way.

One might think the fact that moral truths are *necessarily* true is a reason why it's wrong to take moral uncertainty into account using an analogue of expected utility theory. Under empirical uncertainty, one knows that there is some chance of one outcome, and some chance of another

for example, imprecise credences over moral theories, then we would need to depart from *maximize expected choiceworthiness*. However, our key argument in this chapter is that we should treat moral and empirical uncertainty analogously. So, if we try to accommodate imprecise credences over moral theories, the way in which we should depart from *maximize expected choiceworthiness* should mimic the way in which we should depart from expected utility theory more generally once we allow imprecise credences.

<sup>14</sup> The (risk-neutral) expected value of something (its 'expectation') is just the average of its value in the different cases under consideration weighted by the probability of each case. So the expected choiceworthiness of an option is the average of its choiceworthiness according to the different theories, weighted by the credence in those theories.

<sup>15</sup> One might claim, following Lara Buchak (*Risk and Rationality*, Oxford: Oxford University Press, 2013), that one ought, in general, to endorse a form of risk-weighted expected utility theory. We are perfectly open to this. Our primary claim is that one should endorse maximizing risk-weighted choiceworthiness if and only if risk-weighted expected utility theory is the correct way to accommodate empirical uncertainty. We don't wish to enter into this debate, so for clarity of exposition we assume that the risk-neutral version of expected utility theory is the correct formal framework for accommodating empirical uncertainty.

<sup>16</sup> For an argument of this sort, see Sepielli, "Along an Imperfectly Lighted Path".

outcome. But it doesn't make sense to speak of chances of different moral theories being true (apart from probabilities 1 or 0). And that, one might think, makes an important difference.

However, consider mathematical uncertainty. It is necessarily true whether or not the 1000th digit of the decimal expansion of  $\pi$  is a 7. But, unless we've sat down and worked out what the 1000th digit of  $\pi$  is, we should be uncertain about whether it's 7 or not. And when we need to take actions based on that uncertainty, expected utility theory seems to be the right account. Suppose that one is offered a bet that pays out \$1 if the 1000th digit of  $\pi$  is a 7. How much should one be willing to pay to take that bet? Since there are ten possibilities and the limiting relative frequency of each of them in the decimal expansion of  $\pi$  is equal, it seems one's subjective credence that the 1000th digit of  $\pi$  is a 7 should be 0.1. If so, then, according to expected utility theory, one should be willing to pay 10 cents to take that bet (assuming that, over this range, money doesn't have diminishing marginal value). And that seems exactly right. Even if there's some, highly ideal, sense in which one ought to be certain of all mathematical truths, and act on that certainty, there's clearly a sense of 'ought' which is relative to real-life decision-makers' more impoverished epistemic situation; for that sense of 'ought', expected utility theory seems like the right account of how to make decisions in light of uncertainty. And if this is true in the case of mathematical uncertainty, then the same considerations apply in the case of moral uncertainty as well.<sup>17</sup>

This analogy between decision-making under empirical uncertainty and decision-making under moral uncertainty becomes considerably stronger when we consider that the decision-maker might not even *know* the nature of her uncertainty. Suppose, for example, that Sophie is deciding whether to eat chicken. She's certain that she ought not to eat an animal if that animal is a person, but she is uncertain about whether chickens are persons or not. And suppose that she has no idea whether her uncertainty stems from empirical uncertainty, about chickens' capacity for certain experiences, or from moral uncertainty, about what the sorts of attributes qualify one as a person in the morally relevant sense.

<sup>17</sup> Of course, this means departing from standard probability theory, which assigns probability 1 to all necessary propositions. How to create a formal theory of probability that can reject this idea is a problem that we will leave for another time; however, the fact that we *are* uncertain, and seem justifiably uncertain, in some necessary truths, means that we have to overcome this problem no matter what our view on moral uncertainty. See Michael G. Titelbaum, *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief*, Oxford: Oxford University Press, 2012.

It doesn't seem plausible to suppose that the nature of her uncertainty could make a difference as to what she should decide. It seems even less plausible to think that it could be extremely important for Sophie to find out the nature of her uncertainty before making her decision. But if we think that moral and empirical uncertainty should be treated in different ways, then this is what we're committed to. If her uncertainty stems from empirical uncertainty, then that uncertainty should be taken into account, and everyone would agree that she ought not (in the subjective sense of 'ought') to eat the chicken. If her uncertainty stems from moral uncertainty and moral and empirical uncertainty should be treated differently, then it might be that she should eat the chicken. But then, because finding out the nature of her uncertainty could potentially completely change her decision, she should potentially invest significant resources into finding out what the nature of her uncertainty is. This seems bizarre.

So, as well as pointing out the problems with alternative views, as we did in sections I–II, there seems to be a strong direct argument for the view that moral and empirical uncertainty should be treated in the same way. Under empirical uncertainty, expected utility theory is the standard formal framework. So we should take that as the default correct formal framework under moral uncertainty as well, and endorse maximizing expected choiceworthiness.<sup>18</sup>

#### IV. Two Objections to MEC

In this section we discuss two objections to MEC: that the view is *too demanding* and that it cannot handle the idea that some options are *supererogatory*.

<sup>18</sup> An argument for the risk-neutral version of MEC, in particular, could be made using the non-standard axiomatization of expected utility theory in Martin Peterson, 'From Outcomes to Acts: A Non-Standard Axiomatization of the Expected Utility Principle', *Journal of Philosophical Logic*, vol. 33, no. 4 (August 2004), pp. 361–78. Unlike standard axiomatizations (e.g. John Von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press, 1953), which are given over *lotteries*, Peterson's is given over *outcomes*. This requires an independently motivated interval-scale structure of utility for outcomes, which is usually considered a problem. However, the analogue of utility of outcomes in our case is the choiceworthiness of options, according to a given theory, and we are already supposing this to be at least roughly interval-scale measurable and comparable between theories; so we are in a good position to use this axiomatization to argue for risk-neutral MEC. See also Stefan Riedener, 'Maximising Expected Value under Axiological Uncertainty', BPhil thesis, University of Oxford, 2013 for an axiomatic argument in support of maximizing expected value under evaluative uncertainty.

## Demandingness

The first objection we'll consider is that MEC is too *demanding*: it has implications that require too great a personal sacrifice from us.<sup>19</sup> For example, Peter Singer has argued that members of affluent countries are obligated to give a large proportion of their income to those living in extreme poverty, and that failing to do so is as wrong, morally, as walking past a drowning child whose life one easily could save.<sup>20</sup> Many people who have heard the argument don't believe it to be sound; but even those who reject the argument should have at least some credence in its conclusion being true. And everyone agrees that it's at least permissible to donate the money. So isn't there a dominance argument for giving to fight extreme poverty? The decision situation seems to be as in Table 2.7.

If so, then it is appropriate for us, as citizens of affluent countries, to give a large proportion of our income to fight poverty in the developing world. But (the objection goes) that is too much to demand of us. So Dominance, and therefore MEC, should be rejected.

Our first response to this objection is that it is guilty of double-counting.<sup>21</sup> Considerations relating to demandingness *are* relevant to consideration of what it is appropriate to do under moral uncertainty. But they are relevant because they are relevant to what credences one ought to have across

**Table 2.7**

	Singer's conclusion is correct	Singer's conclusion is incorrect
Give	Permissible	Permissible
Don't Give	Impermissible	Permissible

<sup>19</sup> Weatherson hints at this objection in 'Review of Ted Lockhart, *Moral Uncertainty and Its Consequences*'; it is made at length in Christian Barry and Patrick Tomlin, 'Moral Uncertainty and Permissibility: Evaluating Option Sets', *Canadian Journal of Philosophy*, vol. 46, no. 6 (2016), pp. 898–923. For discussion, see Sepielli, "Along an Imperfectly Lighted Path", pp. 103–5.

<sup>20</sup> Peter Singer, 'Famine, Affluence, and Morality', *Philosophy & Public Affairs*, vol. 1, no. 3 (Spring 1972), pp. 229–43.

<sup>21</sup> For a response to this objection, see Christian Tarsney, 'Rejecting Supererogationism', *Pacific Philosophical Quarterly*, vol. 100, no. 2 (June 2019), pp. 599–623, sect. 4. <https://doi.org/10.1111/papq.12239>. A separate, more deflationary, response would be to re-emphasize that we are not talking about permissibility under moral uncertainty, only about what the appropriateness ordering is, and to contend that demandingness is about what options are permissible under moral uncertainty. However, we think that there are interesting issues here, so we will assume that our objector finds even the fact that certain very self-sacrificial actions to be more appropriate than all other options to be implausibly demanding.

different moral theories. If they were *also* taken to be relevant to which theory of decision-making under moral uncertainty is true, then one has given demandingness considerations more weight than they should have. Consider an analogy: it would clearly be incorrect to argue against MEC because, in some cases, it claims that it is appropriate for one to refrain from eating meat, even though (so the objection goes) there's nothing wrong with eating meat. That would be double-counting the arguments against the view that it is impermissible to eat meat; in general, it seems illegitimate to move from claims about first-order moral theories to conclusions about which theory of decision-making under moral uncertainty is true.

However, we do think that it's reasonable to be suspicious of this dominance argument for giving a large proportion of one's income to fight global poverty. We think that a theory of decision-making under moral uncertainty should take into account uncertainty about what the *all-things-considered* choice-worthiness ordering is. And the decision-maker who rejects Singer's argument should have some credence in the view that, all things considered, the most choiceworthy option is to spend the money on herself (or on her family and friends). This would be true on the view according to which there is no moral reason to give, whereas there is a prudential reason to spend the money on herself (and on her friends). So the decision-situation for a typical decision-maker might look as in Table 2.8.

Given this, what it's appropriate to do depends on exactly how likely the decision-maker finds Singer's view. It costs approximately \$3,200 to save the life of a child living in extreme poverty,<sup>22</sup> and it would clearly be wrong, on the common-sense view, for someone living in an affluent country not to

**Table 2.8**

	Singer's argument is correct	Singer's argument is mistaken + prudential reasons to benefit oneself	Singer's argument is mistaken + no prudential reasons to benefit oneself
Give	Permissible	Slightly wrong	Permissible
Don't Give	Gravely wrong	Permissible	Permissible

<sup>22</sup> GiveWell, 'Against Malaria Foundation.' Note that GiveWell's estimated cost per young life saved-*equivalent* is about \$3,200. That is, GiveWell estimates that, if you give \$3,200 to the Against Malaria Foundation, you will in expectation cause an outcome that, according to the values of the median GiveWell staff member, is morally equivalent to saving the life of one young child. For discussion, see Ajeya Cotra, 'AMF and Population Ethics,' The GiveWell Blog, 12 December 2016. <http://blog.givewell.org/2016/12/12/amf-population-ethics/>

save a drowning child even if it were at a personal cost of \$3,200. It seems to us that this intuition still holds even if it cost \$3,200 to prevent a one in ten chance of a child drowning. In which case, the difference in choiceworthiness between giving and not-giving, given that Singer's conclusion is true, is at least ten times as great as the difference in choiceworthiness between giving and not-giving, given that Singer's conclusion is false. So if one has at least 0.1 credence in Singer's view, then it would be inappropriate not to give. However, the intuition becomes much more shaky if the \$3,200 only gave the drowning child an additional one in a hundred chance of living. So perhaps the difference in choiceworthiness between giving and not-giving, on the assumption that Singer's conclusion is true, is less than one hundred times as great as the difference in choiceworthiness between not-giving and giving, on the assumption that Singer's conclusion is false. In which case, it would be appropriate to spend the money on oneself if one has less than 1% credence that Singer's conclusion is true.

The above argument was very rough. But it at least shows that there is no two-line knockdown argument from moral uncertainty to the appropriateness of giving. Making that argument requires doing first-order moral philosophy, in order to determine how great a credence one should assign to the conclusion of Singer's view. And that, we think, should make us a lot less suspicious of MEC. The two-line argument seemed too easy to be sound. For example, Weatherson commented that: "The principle has some rather striking consequences, so striking we might fear for its refutation by a quick *modus tollens*"<sup>23</sup> and

I'm arguing against philosophers who, like Pascal, think they can convince us to act as if they are right as soon as we agree there is a non-zero chance that they are right. I'm as a rule deeply sceptical of any such move, whether it be in ethics, theology, or anywhere else.<sup>24</sup>

We agree with him on these comments. But the error was not with MEC itself: the error was that MEC was being applied in too simple-minded a way.<sup>25</sup> We shall come back to the question of the practical implications of moral uncertainty in much more detail in Chapter 8.

<sup>23</sup> Weatherson, 'Review of Ted Lockhart, *Moral Uncertainty and Its Consequences*', p. 694.

<sup>24</sup> Weatherson, 'Running Risks Morally', p. 145.

<sup>25</sup> We think this reply is also effective against Barry and Tomlin, 'Moral Uncertainty and Permissibility: Evaluating Option Sets'. Barry and Tomlin present an alternative account, which is supposed to avoid the demandingness objection. However, it suffers from some significant

## Supererogation

The second objection we'll consider is that MEC cannot properly accommodate the fact that theories include the idea of *supererogation*. That is, two options might both be permissible, but one may be, in some sense, morally superior to the other. Insofar as MEC is sensitive *only* to a theory's choiceworthiness function, and permissibility is defined as optimal choiceworthiness, it may seem to neglect this aspect of morality.<sup>26</sup>

In order to determine whether this is a good objection to MEC, we need to understand what supererogation is. Accounts of supererogation can be divided into three classes.<sup>27</sup>

The first and most popular type of account is the Reasons Plus type of account. On this type of account, the normative status of an option (in particular, whether it is obligatory or merely supererogatory) is determined by both the choiceworthiness of the option, and by some other factor, such as praiseworthiness.<sup>28</sup>

According to one account, for example, an option is permissible iff it's maximally choiceworthy; an option is supererogatory if it's permissible and if choosing that option is praiseworthy.

On this account, MEC has little trouble with supererogation. Different theories might label some options as supererogatory because of the reactive attitudes that it is appropriate for others to have towards people who choose those options. But that doesn't change the theory's choiceworthiness functions; so it doesn't affect how MEC should treat different theories.

If this account of supererogation were true, it would be true that there are elements of morality on which MEC is silent. If one regards praiseworthiness

unclearly. Moreover, it requires us to make sense of normative assessments of *sets* of options as well as *second-order* moral evaluations: it is morally bad that a moral theory is morally demanding. We find both of these requirements problematic.

<sup>26</sup> See Ted Lockhart, *Moral Uncertainty and Its Consequences*, New York: Oxford University Press, 2000.

<sup>27</sup> We take this classification, and the references below, from Sepielli, "Along an Imperfectly Lighted Path", pp. 238–45.

<sup>28</sup> For examples of this type of account, see Joseph Raz, 'Permissions and Supererogation', *American Philosophical Quarterly*, vol. 12, no. 2 (April, 1975), pp. 161–8; Bernard Williams, 'Persons, Character, and Morality', in Amélie Oksenberg Rorty (ed.), *The Identities of Persons*, Los Angeles: University of California Press, 1977, pp. 197–216; and Susan Wolf, 'Moral Obligations and Social Commands', in Samuel Newlands and Larry M. Jorgensen (eds), *Metaphysics and the Good: Themes from the Philosophy of Robert Merrihew Adams*, Oxford: Oxford University Press, 2009, pp. 343–67.



and blameworthiness as important moral concepts, then one might wish to extend our account: one might wish to develop an account of when one is blameworthy when acting under moral uncertainty, in addition to an account of what one ought to do under moral uncertainty. This is a major topic that we put aside in the book. But it doesn't pose a problem for MEC itself.

The second type of account of supererogation we may call the Kinds of Reasons accounts. On these accounts, options with the same level of choiceworthiness gain different normative statuses in virtue of their position in some other ordering.<sup>29</sup>

According to one possible account, for example, an option is permissible iff it's all-things considered maximally choiceworthy; an option is supererogatory iff it's all-things-considered maximally choiceworthy and better in terms of other-regarding reasons (rather than prudential or esthetic reasons) than all other maximally choiceworthy options.

On this account, again, there seems to be little that is problematic for MEC, since it is a function from the all-things-considered choiceworthiness functions to an appropriateness ordering. Within this theory, we can accept that some maximally choiceworthy actions can be better in terms of other-regarding reasons than others.

The third type of account of supererogation we may call Strength of Reasons accounts. On this view, an option is obligatory iff it's maximally choiceworthy and the reasons in favour of it are sufficiently strong compared to other available options (that is, if the maximally choiceworthy option is only a little more choiceworthy than the other permissible options, in some sense of 'only a little' that would need to be defined).

This account poses some problems for MEC because, on this account, there is more reason to choose one option *x* than another option *y* even though both options are permissible. This leaves us with a decision. Are both options maximally choiceworthy (because both are maximally permissible)? Or is the one we have more reason to choose more choiceworthy?

We don't find this view particularly plausible. However, we suggest that, if you endorse such an account, you should regard option *A* as more choiceworthy than option *B* even if both options are permissible. If you

<sup>29</sup> For examples of this type of account, see Douglas W. Portmore, 'Position-Relative Consequentialism, Agent-Centered Options, and Supererogation', *Ethics*, vol. 113, no. 2 (January 2003), pp. 303–32; Michael Zimmerman, *The Concept of Moral Obligation*, Cambridge: Cambridge University Press, 1996; and Ruth Chang, 'Voluntarist Reasons and the Sources of Normativity', in David Sobel and Steven Wall (eds), *Reasons for Action*, Cambridge: Cambridge University Press, 2009, pp. 243–71.

were to endorse such a view, then you might wish to have a separate theory of how to aggregate deontic statuses under moral uncertainty; what it is rationally permissible to do under moral uncertainty might come apart from what the most appropriate option is.<sup>30</sup> However, we do not attempt that project here; our project is just about the strengths of reasons that we have, on different theories, and how to aggregate them in conditions of uncertainty.

## Conclusion

In this chapter we have argued that, in conditions of interval-scale measurable and intertheoretically comparable choiceworthiness, moral and empirical uncertainty should be treated in the same way. Because we take expected utility theory to provide the default formal framework for taking empirical uncertainty into account, that means we think that *maximize expected choiceworthiness* is the default account for making decisions in the face of moral uncertainty. In the next chapter, we will discuss what the right theory is when moral theories are incomparable and provide merely ordinal choiceworthiness.

<sup>30</sup> To see that this is so, consider a decision-maker who is certain in a moral view on which this view of supererogation is correct. If one thought that only appropriate options were rationally permissible, then there would be situations in which the decision-maker would be certain that two options were morally permissible, but where only one option was rationally permissible (in the sense of rational permissibility that is relevant to decision-making under moral uncertainty). This seems problematic. We thank Christian Tarsney for this point.