

## *Why Maximize Expected Choice-Worthiness?*<sup>1</sup>

WILLIAM MACASKILL AND TOBY ORD  
University of Oxford

This paper argues in favor of a particular account of decision-making under normative uncertainty: that, when it is possible to do so, one should *maximize expected choice-worthiness*. Though this position has been often 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 favor. After dealing with some preliminaries and giving the basic motivation for taking normative uncertainty into account in our decision-making, we consider and provide new arguments against two rival accounts that have been offered—the accounts that we call ‘My Favorite Theory’ and ‘My Favorite Option’. We then give a novel argument for *comparativism*—the view that, under normative uncertainty, one should take into account both probabilities of different theories and magnitudes of choice-worthiness. Finally, we further argue in favor of maximizing expected choice-worthiness and consider and respond to five objections.

### Introduction

Normative uncertainty is a fact of life. Suppose that Michael has £20 to spend. With that money, he could eat out at a nice restaurant. Alternatively, he could eat at home and pay for four long-lasting insecticide-treated bed nets that would protect eight children against malaria. Let’s suppose that Michael knows all the morally relevant empirical facts about what that £20 could do. Even so, it might be that he still doesn’t know whether he’s obligated to donate that money or whether it’s permissible for him to pay for the meal out, because he just doesn’t know how strong his moral obligations to distant strangers are. If so, then even though Michael knows all the relevant empirical facts, he doesn’t know what he ought to do.

Or suppose that the members of a government are making a decision about whether to tax carbon emissions. Let’s assume that they know all the relevant facts about what would happen as a result of the tax: it would make presently existing people worse off, since they would consume less oil and coal, and would therefore be less economically productive; but it would slow the onset of climate change, thereby increasing the welfare of people living in the future. But the members of the government don’t know how to weigh the interests of future people against the interests of presently existing people. So, again, those in this government don’t ultimately know what they ought to do.

In both of these cases, the uncertainty in question is not uncertainty about what will happen, but rather is fundamental normative uncertainty. Recently, some

philosophers have suggested that there are norms that govern how one ought to act which take into account one's fundamental normative uncertainty.<sup>2</sup>

Our primary aim in this article is to argue in favor of a particular partial account of decision-making under normative uncertainty: that, when all moral views under consideration give an interval-scale measurable and intertheoretically comparable notion of choice-worthiness, one should *maximize expected choice-worthiness*.<sup>3</sup>

Though this position has been often 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 favor. The structure of our argument is as follows. After dealing with some preliminaries in section 1, and giving the basic motivation for taking normative uncertainty into account in our decision-making in section 2, we devote sections 3 and 4 to considering and providing new arguments against two rival accounts that have been proposed in the literature—the accounts which we call 'My Favorite Theory' and 'My Favorite Option'. In section 5, we give a novel argument for *comparativism*—the view that, under normative uncertainty, one should take into account both probabilities of different theories and magnitudes of choice-worthiness. In section 6, we further argue in favor of maximizing expected choice-worthiness (MEC). In section 7, we consider and respond to five objections.

## 1. Preliminaries

Over the centuries, many different normative theories have been proposed and these have come in many different forms. In order to consider how to make decisions when one is uncertain about what one has reason to do, we require a framework that is generic enough to encompass as many of these theories as possible. At the same time, it must be specific enough to get any traction on the problem, to allow us to make points clearly, and to avoid unnecessary verbiage.

This article is focused on decision-making under normative uncertainty. By this we mean uncertainty over what one has all-things-considered reason to do: this uncertainty therefore covers both moral uncertainty and prudential uncertainty, though our focus will be on moral uncertainty. Our article is not concerned with uncertainty over epistemology or over which decision theory or theory of rationality is correct; we believe these raise interesting questions, but they must be left for another time.<sup>4</sup>

We make the following structural assumptions about what a decision under normative uncertainty looks like. We consider a decision-maker choosing from a set of exhaustive and mutually exclusive options ( $A, B, C, \dots$ ). These options could be acts, or they could be plans of action, or anything else that could be the subject of choice and normative assessment.

We suppose that the decision maker has credence in each of a set of first-order normative theories ( $T_1, T_2, T_3, \dots$ ). We will normally talk about these theories as if they are complete stand-alone normative theories, such as a particular form of utilitarianism. However, they could often just as well represent partially specified theories, or particular normative considerations regarding the options at hand, such as whether killing is equivalent to letting die.

We will sometimes represent the credence in a given theory with a real number between zero and one, or as a percentage. This is not to assume that we have precise credences in these theories. Nothing will turn upon the exact values of these credences, and we believe that a modified version of the view we defend could just as well be defended if we were to use imprecise credences.<sup>5</sup>

We will assume that the theories under consideration assess these options in terms of *choice-worthiness*, which represents the strength of the reasons for choosing a given option. This need not be quantitative: it could just provide an ordering as to which options are more choice-worthy than others. Moreover, this relation need not hold between all pairs of options: some might be incomparable in terms of choice-worthiness. In this article, however, we in general consider only theories that can make quantitative judgments about choice-worthiness, such that one option might be slightly more choice-worthy than a second, but much more choice-worthy than a third. We also only consider situations where one can make quantitative comparisons of choice-worthiness *across* theories. If these conditions were not met, then the idea of *expected choice-worthiness* would be undefined, and *maximize expected choice-worthiness* would be straightforwardly inapplicable. In other work, we develop accounts of what to do under normative uncertainty when these conditions are not met,<sup>6</sup> and what to do in mixed situations, where these conditions are met by some theories and not by others.<sup>7</sup>

We will occasionally use numbers to represent levels of choice-worthiness. For the time being, we will say that an option is *permissible* (right) iff it is maximally choice-worthy, and *impermissible* (wrong) if it is not maximally choice-worthy; we will return to the relationship between choice-worthiness and deontic status in section 7, iii, where we discuss how to handle supererogation.

Some decisions made under normative uncertainty are intuitively superior to others. For example, intuitively there is something important to be said against choosing option *A* when all theories in which you have credence consider it to be impermissible, and they all consider option *B* to be permissible—even if, according to the true normative theory, action *A* is the morally correct choice. We shall use the term *appropriate* to make such assessments of options under normative uncertainty, where an appropriate action is what would be selected by a rational and morally conscientious agent who had the same set of options and beliefs. As we use the term, more than one option could be appropriate, some options may be more appropriate than others, some may be incomparable, and there may be degrees of appropriateness.

To keep things simple, we shall only consider examples in which there is no descriptive uncertainty, though it is of course possible to simultaneously have descriptive and moral uncertainty.

## 2. The motivation for taking normative uncertainty into account

There are two main motivations for taking normative uncertainty into account in our decision-making. The first is simply an appeal to intuitions about cases. 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 pleasant, so she has no prudential reason for preferring one over the other. Let's suppose that, according to the true normative theory, both of these options are equally choice-worthy: animal welfare is not of moral value so there is no moral reason for choosing one meal over another. Let's suppose that Jane has high credence in that view. But she also finds plausible the view that animal welfare is of moral value, according to which the risotto is the more choice-worthy option.

In this situation, choosing the risotto over the foie gras is more choice-worthy according to some theories and less choice-worthy according to none. In the language of decision theory, the risotto *dominates* the foie gras. So it seems very clear that, in some sense at least, Jane would act incorrectly if she were to choose the foie gras. But, if so, then there must be norms that take into account Jane's first-order normative uncertainty.

A second motivation for taking normative uncertainty into account is based on the idea of action-guidingness. There has been a debate concerning whether there are norms that are relative to the decision-maker's beliefs or credences ('subjective' norms), either in addition to norms that are not relative to the decision-maker's beliefs or credences ('objective' norms).<sup>8</sup> The principal argument for thinking that there must be subjective norms is that objective norms are not sufficiently action-guiding. Consider the following case:<sup>9</sup>

*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 each 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, 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.

Her decision can be represented in the following table:

	Greg has condition A – 50%	Greg has condition C – 50%
A	Completely cured	Dead
B	Almost completely cured	Almost completely cured
C	Dead	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?

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 Y is low, it seems that it would be reckless for Susan to administer drug C. As far as she knows, in doing so she'd 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.

In this case, the objective consequentialist's recommendation—"do what actually has the best consequences"—is not useful advice for Susan. It is not a piece of advice that she can act on, because she is uncertain about what action actually has the best consequences (and about what process would reduce that uncertainty). So one might worry that the objective consequentialist's recommendation is not sufficiently action-guiding. In contrast, so the argument goes, if there are subjective norms, then the decision-maker will very often know what these norms require. So the thought that there should exist sufficiently action-guiding norms motivates the idea that there exist subjective norms.

Similar considerations motivate the idea that there are norms that are relative to normative 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, in which case sufficiently action-guiding norms must take into account normative uncertainty as well. Consider the following variant of the case:<sup>10</sup>

### *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. She has a vial of a drug that can help. If she administers all of the drug to Anne, she will survive but with disability, at half the level of welfare she'd have if healthy. If she administers all of the drug to Charlotte, she will be completely cured. If she 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 healthy. She 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 the following table, using numbers to represent how good each outcome would be for each individual.

	Anne's welfare	Charlotte's welfare
A	50	0
B	49	49
C	0	100

While the theories on whether chimpanzee welfare counts disagree strongly about the relative value of options A and C, they both hold that option B is only slightly inferior to the best option. This can be represented as follows:

	Chimpanzee welfare is of no moral value – 50%	Chimpanzee welfare is of full moral value – 50%
A	permissible	extremely wrong
B	slightly wrong	slightly wrong
C	extremely wrong	permissible

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?

In the first variant of this case, 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. Moreover, ‘do what’s right given your empirical credences’ isn’t useful advice for Susan, because it doesn’t take into account her uncertainty over first-order normative views. So, it seems that, in this case, norms that take into account only Susan’s empirical uncertainty aren’t sufficiently action-guiding. And this motivates the idea that there are norms that take into account both one’s empirical uncertainty and one’s normative uncertainty.

For these reasons, we think that there is a clear *prima facie* case in favor of the idea that there are norms that take first-order normative uncertainty into account. In this article, we do not argue further for this idea. Our main intention is to explore this idea, rather than defend it at length.<sup>11</sup>

### 3. Against My Favorite Theory

One might think that, under normative uncertainty, one should simply follow the normative 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:

[T]he proper approach to uncertainty about the rightness of ethical theories is to determine the one most likely to be right, and to act in accord with its dictates.<sup>12</sup>

It has also recently been defended by Gustafsson and Torpman.<sup>13</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  $S$  has highest credence in.

This is an elegant 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 normative 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.

Before addressing the first problem, we turn to the second fixable problem, which is that the view violates the following principle:

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

MFT violates this in the following case:<sup>14</sup>

	$T_1 - 40\%$	$T_2 - 60\%$
A	permissible	permissible
B	impermissible	permissible

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. 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, at a cost of sacrificing much of the view's elegance, by modifying it slightly. This is what Gustafsson and Torpman do. 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 normative theory  $T_i$  such that
  - (a)  $T_i$  is in the set  $\mathcal{T}'$  of theories that are at least as credible as any theory for  $S$  and
  - (b)  $S$  has not violated  $T_i$  more recently than any other theory in  $\mathcal{T}'$ , and
- (2) There is no option  $B$  and no normative theory  $T_j$  such that
  - (a)  $T_j$  requires  $B$  and not  $A$  and
  - (b) No 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 one does not perform bizarre courses of action: 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 is that it's going to run into what we'll call the *problem of theory-individuation*. 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 follows:

	Non-consequentialism – 40%	Utilitarianism – 60%
A	impermissible	permissible
B	permissible	impermissible

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. So 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 follows:

	Non-Consequentialism – 40%	Utilitarianism <sub>1</sub> – 30%	Utilitarianism <sub>2</sub> – 30%
A	impermissible	permissible	permissible
B	permissible	impermissible	impermissible

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 normative theories:

Regard normative theories  $T_i$  and  $T_j$  as versions of the same 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>15</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 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 follows:

	$U - 2\%$	$P_1 - 1\%$	$P_2 - 1\%$	$\dots$	$P_{98} - 1\%$
A	permissible	impermissible	impermissible	$\dots$	impermissible
B	impermissible	permissible	permissible	$\dots$	permissible

On Gustafsson and Torpman's version of MFT, the appropriate option for Tracy is A, even though she's almost certain that it's wrong to choose A. This can't be right.<sup>16</sup>

#### 4. Against My Favorite Option

The solution to these last two problems with My Favorite Theory 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>17</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 has serious problems of its own.

The first problem is that it is susceptible to money-pumps: that is, it can recommend that the decision-maker makes a series of decisions such that she is predictably

left strictly worse off than she was when she started. Consider the following case. Ursula has equal credence in three consequentialist theories. These theories value the options A, B, C, and A' as follows (with higher numbers representing a more choice-worthy option):

	$T_1 - 33\%$	$T_2 - 33\%$	$T_3 - 34\%$
A	4	2	3
B	1	3	4
C	2	4	1
A'	3	1	2

Suppose that Ursula has recently chosen option A. But now she is given the opportunity to change her mind. She is offered the choice between A and B. This decision-situation is represented by rows one and two above. A is permissible according to  $T_1$  and impermissible according to  $T_2$  and  $T_3$ ; B is permissible according to both  $T_2$  and  $T_3$  and impermissible according to  $T_1$ . B has the highest probability of permissibility, so is the most appropriate option according to MFO. Ursula, following MFO, chooses B.

Ursula is then offered the choice between B and C. This second decision-situation is represented by rows two and three. C is permissible according to two of the theories, while B is only permissible according to one, so C is the option with the highest probability of permissibility, and thus the most appropriate option according to MFO. So Ursula chooses C.

Next, however, let us suppose that Ursula is offered the choice between C and A'. This third decision-situation is represented by the final two rows. In this decision-situation, C is only permissible according to one theory, while A' is permissible according to two, so A' is the option with the highest probability of permissibility, and thus the most appropriate option according to MFO. So Ursula chooses A'.

But Ursula is *certain* that A' is less choice-worthy than A. By following MFO, Ursula has ended up with an option that is inferior to the one she started with. Moreover, the process could be iterated. Ursula could start off with an extremely valuable option A, but then be led to choose worse and worse options, A', A'', etc. until eventually she ends up with an option that is very bad indeed. This is known as a money-pump (or 'value pump') and it is a problem for MFO.<sup>18</sup>

This problem arose because MFO only considers the ordering of choice-worthiness between options and ignores information about the degree of choice-worthiness. In the next section we shall see that this produces even greater problems.

### 5. In favor of comparativism

Let's define *comparativism* as the view that what it's appropriate to do depends upon both the credences that the decision-maker assigns to different moral theories and the degrees of choice-worthiness that those theories assign to different options.

MFT and MFO are both non-comparativist: they are sensitive only to the credences that the decision-maker has across moral theories.

But we can construct clear counterexamples to non-comparativism. First, suppose that your credence is split between two theories, with the second theory being 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 both on the right act and on the magnitude of what is at stake. The 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 with the following table:

	$T_1 - 51\%$	$T_2 - 49\%$
A	permissible	gravely wrong
B	slightly wrong	permissible

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 the second variant of *Susan and the Medicine*:

*Susan and the Medicine - II*

	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

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 account of decision-making under moral uncertainty is false if it says the appropriate option is always the maximally choice-worthy option according to some theory in which one has credence. This includes MFT, MFO, and their variants.

Moreover, this case shows that one understanding of the central question for decision-making under normative uncertainty that has been given in the literature is wrong. Jacob Ross suggests that the central question is “what ethical theories are worthy of acceptance and what ethical theories should be rejected,” where Ross defines acceptance as follows:<sup>19</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 normative 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>20</sup>

### **6. In favor of treating moral and empirical uncertainty analogously**

In the previous section we saw an argument in favor of comparativism: that the appropriate option to choose depends upon both the decision-maker’s credences over theories, and the degrees of choice-worthiness that those theories assign to options. But which form of comparativism is correct? In this section we argue that, when choice-worthiness differences are comparable across theories, we should handle normative 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. So maximizing expected choice-worthiness should be the standard account of how to handle moral uncertainty.<sup>21</sup>

We are perfectly open to those who wish to follow Lara Buchak<sup>22</sup> and endorse a form of risk-weighted expected utility theory, or some other departure from expected utility theory. Our primary claim is that one should endorse maximizing risk-weighted choice-worthiness 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.

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

*Maximize Expected Choice-worthiness (MEC)*: when we can compare degrees of choice-worthiness between theories in which we have credence, *A* is an appropriate option iff *A* has the maximal expected choice-worthiness.

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>23</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 decision-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 that the default view is that moral and empirical uncertainty should be treated in the same way.

One might think that the fact that moral truths are *necessarily* true is a reason why it's wrong to take normative 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 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 1000<sup>th</sup> digit of  $\pi$  is a 7. But, unless we've sat down and worked out what the 1000<sup>th</sup> digit of  $\pi$  is, we should be uncertain about whether it's a 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 \$10 if the 1000<sup>th</sup> 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 1000<sup>th</sup> digit of  $\pi$  is a 7 should be 0.1. If so, then, according to expected utility theory, one should be willing to pay \$1 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. And if this is true in the case of mathematical uncertainty, then the same considerations apply in the case of normative uncertainty as well.

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.

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 normative 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 normative 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 3–5, 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 maximising expected choice-worthiness.

## 7. Objections and responses

### *i. Disanalogy with Prudential Reasoning*

One line of objection to MEC comes from Brian Weatherson<sup>24</sup> (though he uses the argument against there being any norms that govern decision-making under normative uncertainty at all, rather than MEC in particular). The objection is that MEC doesn't seem plausible in cases where there are only prudential reasons at stake. 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, shows that Dominance is false, and therefore shows that MEC is false.

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: the time that Bob would spend in the art gallery, having an experience that 'leaves him cold', could presumably be doing something else more enjoyable instead.<sup>25</sup> In which case, depending on how exactly the example was specified, MEC 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. But route B passes by great works of architecture than Bob hasn't seen before, whereas route A does not. In this case, where there is some probability that art has prudential value, MEC 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 MEC gets the right answer in purely prudential cases. Consider the following case.

*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 in to 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 achieve 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 was 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 that 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 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, further undermining Weatherston's objection.<sup>26</sup> While it can be a bit harder to come up with clear examples concerning prudence (due to the much greater agreement between people on what constitutes a good life than on how to act morally), we don't see any difference in the force of the arguments in favor of MEC whether we're considering moral reasons or prudential reasons.

*ii. Demandingness?*

A second objection is that MEC is too *demanding*. It has implications that require too great a personal sacrifice from us.<sup>27</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>28</sup> Many people who have heard the argument don't believe it to be sound; but even those who reject the argument should give at least some credence that its conclusion is 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? That is, doesn't the decision situation look like this?:

---

	Singer's conclusion is correct	Singer's conclusion is incorrect
Give	permissible	permissible
Don't Give	impermissible	permissible

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, so 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. Considerations relating to demandingness *are* relevant to 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 different moral theories. If they were *also* relevant to which account of decision-making under normative uncertainty is true, then one has given demandingness considerations twice the weight that they should have. As 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 normative theories to conclusions about which theory of decision-making under normative uncertainty is true. One might instead respond that norms governing decision-making under normative uncertainty are norms about rationality, and argue that the claim that MEC is too rationally demanding is a separate objection from the claim that some first-order normative theories are too morally demanding. But it seems odd to us to claim that a theory of rationality, if that's what MEC is, can be too demanding.

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 an account of decision-making under normative 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 reasonable credence in the view that, all things considered, she ought to spend the money on herself (or on her friends and family). 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 follows:

	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

Given this, what it's wrong to do depends on exactly how likely the decision-maker finds Singer's view. It costs approximately \$3400 to do as much good as saving the life of a child living in extreme poverty,<sup>29</sup> and it would clearly be wrong, on the common sense view, for someone living in an affluent country not to save a drowning child even if it were at a personal cost of \$3400. It seems to us that this intuition still holds even if it cost \$3400 to prevent a one in ten chance of a child drowning. In which case, the difference in choice-worthiness between giving and not-giving, given that Singer's conclusion is true, is at least 10 times as great as the difference in choice-worthiness between giving and not-giving, given that Singer's conclusion is false. So if one has at least 10% credence in Singer's view, then it would be inappropriate not to give. However, the intuition becomes much more shaky if the \$3400 only gave the drowning child an additional one in a hundred chance of living. So perhaps the difference in choice-worthiness between giving and not-giving, on the assumption that Singer's conclusion is true, is less than 100 times as great as the difference in choice-worthiness 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 normative 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.<sup>30</sup> And that's true. But the error was not with MEC itself: the error was that MEC was being applied in too simple-minded a way.

iii. *Supererogation*

The third objection we'll consider is that MEC cannot properly accommodate the fact that some 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 choice-worthiness ordering, it may seem to neglect this aspect of morality.<sup>31</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>32</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 choice-worthiness of the option, and by some other factor, such as praiseworthiness.<sup>33</sup>

According to one account, for example, an option is permissible iff it's maximally choice-worthy; 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 choice-worthiness orderings, 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 that MEC is silent on. If one regards praiseworthiness 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 this article. 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 choice-worthiness gain different normative statuses in virtue of their position in some other ordering.<sup>34</sup>

According to one possible account, for example, an option is permissible iff it's all-things-considered maximally choice-worthy; an option is supererogatory if it's all-things-considered maximally choice-worthy and better in terms of other-regarding reasons (rather than prudential reasons) than all other maximally choice-worthy 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 choice-worthiness orderings to an appropriateness ordering. Within this account, we can accept that some maximally choice-worthy 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 if it's maximally choice-worthy and the reasons in favor of it are sufficiently strong compared to other available options; an option is supererogatory iff it's maximally choice-worthy and the reasons in favor of it are not sufficiently strong compared to other available options (that is, if the maximally choice-worthy option is only a little more choice-worthy than the other permissible options, in some sense of 'only a little' that would need to be defined).

Again, however, this notion of supererogation would not pose a problem for MEC, which is precisely concerned with the strength of reasons one has, on different theories, for choosing different options.

The account of supererogation that would most pose problems for MEC is one on which there is more reason to choose one option *A* than another option *B* even though both options are permissible. This leaves us with a decision. Are both options maximally choice-worthy (because both are permissible)? Or is the one we have more reason to choose more choice-worthy?

We find it very difficult to understand this view, if it is not to be understood in the terms of one of the three accounts given above. However, we suggest that, if you endorse such an account, you should regard option *A* as more choice-worthy than option *B* even if both options are permissible. If you were to endorse such a view, then you might wish to have a separate theory of deontic statuses under normative uncertainty: that many options may be ‘appropriate’ even if, among those appropriate options, some are more appropriate than others. However, we do not attempt that project here.

#### *iv. Intertheoretic Incomparability*

The fourth objection we consider points to the fact that, in order for MEC to be applicable, we must be able to make sense of *magnitudes* of choice-worthiness, and that these magnitudes must be comparable across all theories in which one has credence. We must be able to say, at least roughly, how much greater the difference in choice-worthiness between *A* and *B* is, on  $T_1$ , compared to the difference in choice-worthiness between *C* and *D*, on  $T_2$ .

In response to this point, we acknowledge that sometimes theories do not provide magnitudes of choice-worthiness, and that, even when they do, sometimes magnitudes of choice-worthiness are not comparable across those theories. In those situations, MEC is not applicable. We therefore do not claim that MEC is a perfectly general account. Rather, we defend MEC as the correct account of how to take into account normative uncertainty *only* when magnitudes of choice-worthiness are comparable across theories.<sup>35</sup>

The objector might argue that these conditions are *never* met, because choice-worthiness is never comparable across theories.<sup>36</sup> In which case we have given an account that is never applicable.

It’s true that the question of how to make intertheoretic comparisons is a tricky one. There are several accounts in the literature, but none that we find completely plausible.<sup>37</sup> But what suffices for the purpose of this article is to note that there are clearly *some* cases where choice-worthiness is comparable across theories.<sup>38</sup> The purpose of an account of intertheoretic comparisons is to give a plausible account of the easy cases, where we have clear intuitions about which intertheoretic comparisons hold, in order that we can come to have better view about what intertheoretic comparisons hold, if they do, in the harder cases. But it is unwarranted, simply in the absence of a compelling theory of intertheoretic comparisons, to think that there are no cases in which intertheoretic comparisons hold.

The first class of ‘easy’ cases is where the two theories are very similar. Consider, for example, the view that Singer proposes in ‘Famine, Affluence and Morality’ (recalling that the view that Singer proposes in that article is *not* utilitarianism, but rather is a modification of common-sense ethics). The claim, “It’s more important to give to Oxfam on Singer’s view than it is on the common-sense view” is obviously true. But this is an intertheoretic comparison claim. So to deny that intertheoretic comparisons are ever possible is to say that the above claim is not true. But that seems very implausible. Singer’s view only differs from common-sense morality in one respect: in the importance it assigns to donating to fighting global poverty. So it seems perfectly obvious that Singer’s view and the common-sense view agree on all other aspects of morality, such as on the wrongness of stealing for personal gain. In which case we can easily make an intertheoretic comparison between the two theories, basing that comparison on the shared agreement between the two views.

The second class of ‘easy’ cases is where the two theories disagree only on the extension of bearers of value. Consider (i) utilitarianism, according to which the goodness of an outcome is equal to the sum total of wellbeing across all people and (ii) utilitarianism\*, according to which the goodness of an outcome is equal to the sum total of wellbeing across all people except Richard Nixon. Utilitarianism and utilitarianism\* only differ in the value that they assign to Richard Nixon. So the intertheoretic comparison is perfectly obvious: the two theories agree on the value of every person except Richard Nixon.

It is true that *sometimes* there seems to be no plausible intertheoretic comparison. Comparing average and total utilitarianism might be such a case. In those cases, MEC will not be applicable, and one will have to use an alternative. But our defense of MEC is only intended to be limited to cases where one can make intertheoretic comparisons. And there are clearly some such cases. So MEC is applicable.

#### v. *Fanaticism*

The fourth objection is that MEC will result in *fanaticism*: that is, the expected choice-worthiness will be dominated by theories according to which moral situations are incredibly high stakes.<sup>39</sup> 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 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 0.1. Absolutism agrees that it is choice-worthy to save lives, and that it’s more choice-worthy 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. That is, the decision-situation looks as follows:

	Utilitarianism – 99%	Absolutism – 1%
Lie	+9.9	$-\infty$
Don't lie	0	0

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."<sup>40</sup> 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 contribution to the expected values of one's options swamps that of all other theories."

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,<sup>41</sup> 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 everyone else in the world 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 Her 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 the real numbers (which does not include infinitesimals).<sup>42</sup> But, secondly, 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 result 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 lying is also negative infinity. And this will be true for any decision a real-life decision maker faces. For any option, there will always be some theory in which the decision-maker has non-zero credence according to which that option is infinitely wrong, and some non-zero credence in a theory according to which that option is infinitely right. If some 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.<sup>43</sup> 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 good outcomes or astronomically good outcomes. This is a tricky issue within decision theory and in our view no wholly satisfactory solution has been provided.<sup>44</sup> But it is not a problem that is unique to moral uncertainty. And our aim isn't to defend MEC *per se*; the primary claim in this article is that empirical and moral uncertainty should be treated in the same way. So whatever is the best solution to the fanaticism problem under empirical uncertainty is the best solution to the fanaticism problem under moral uncertainty. This means that this issue is not a distinctive problem for normative uncertainty and can, as such, be set aside for the purposes of this paper.

This is our primary response to the objection. However, there are, we think, two more normative uncertainty-specific things that one can say on this issue. 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 choice-worthy one option is than another. Absolutist theories would always rank any option that involves lying as less choice-worthy 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 telling a lie and doing nothing and the difference in choice-worthiness between doing nothing and saving ten lives.

William MacAskill has explored how to take moral uncertainty into account when faced with merely ordinal theories.<sup>45</sup> He notes that in such conditions it is impossible to use MEC and has argued that we instead should use a variant of the

Borda Rule voting system. If we do this, small credences in absolutist theories do not swamp our decision-making. If we were to use the Borda Rule in *Doug's Lie*, then lying would come out as the most appropriate option.

We think that understanding absolutist theories as merely ordinal is a plausible approach. If so, it shows one way in which MEC is limited. However, this would not be a grave limitation of MEC: we think that absolutist views, and in general views that invoke a lexical ordering of wrongs, are quite implausible, so this restriction would only apply to theories that occupy a small portion of our credence distribution.<sup>46</sup> Note, importantly, that treating *absolutist* views as merely ordinal does not mean that we have to treat non-consequentialist views in general as merely ordinal. We can understand many threshold non-consequentialist theories (which posit side-constraints that may be overridden if doing so will produce a sufficient amount of good) as giving a quantitative measure of choice-worthiness. Perhaps, on one theory, the side-constraint against killing may be overridden if, by killing one person, one can thereby save ten thousand lives. If so, then it seems natural to treat the difference in choice-worthiness between killing and doing nothing on this theory as of the same size as the difference in choice-worthiness between doing nothing and saving ten thousand lives. So MEC could still be used in the face of uncertainty over such non-consequentialist views.

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 value of saving a life, we could hold that they agree on the value 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 an 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 is for Doug to lie.

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 follows:

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

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 99<sup>th</sup> as much credence in Utilitarianism-2 as he has in absolutism, MEC would recommend lying.

Taking into account uncertainty about how to normalise 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.

## 8. Conclusion

In this article we have argued that, in conditions where magnitudes of choice-worthiness can be compared across theories, normative 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, we think that, when intertheoretic choice-worthiness comparisons are possible, Maximize Expected Choice-worthiness is the default account of decision-making under normative uncertainty.

<sup>1</sup> For insightful comments on this work, the authors would like to thank Gustaf Arrhenius, Amanda Askill, Nick Bostrom, John Broome, Krister Bykvist, Johan Gustafsson, Dan Moller, Jacob Ross, Andrew Sepielli, Olle Torpman, the participants of the CRNAP workshop on moral uncertainty at Princeton, and two anonymous reviewers.

<sup>2</sup> For example: (Guerrero 2007; Lockhart 2000; Oddie 1995; Ross 2006; Sepielli 2009). John Broome flags the importance of this idea in (2010), though he does not state his views on it.

<sup>3</sup> For work on how this partial account can function as part of a general account of decision-making under moral uncertainty, see (MacAskill 2014) and (MacAskill, Bykvist & Ord under contract).

<sup>4</sup> Though see (MacAskill 2017) for an exploration of the idea of decision-making given decision-theoretic uncertainty.

<sup>5</sup> One might also deny that we should think about decision-making under moral uncertainty in terms of credences at all, perhaps because real-life decision-makers do not possess even imprecise credences. We leave this possibility to this side in this article.

<sup>6</sup> See (MacAskill 2016).

<sup>7</sup> See (MacAskill 2014) and (MacAskill, Bykvist & Ord under contract).

<sup>8</sup> See, for example, (Jackson 1991; Zimmerman 2006; Graham 2010).

<sup>9</sup> This is a ‘Jackson case’, from (Jackson 1991, 462–3). (Regan 1980, 264–5) gives a similar example.

<sup>10</sup> This is a Jackson case under moral uncertainty. A similar case is given by (Zimmerman 2008, 35).

<sup>11</sup> (Weatherson 2014) and (Harman 2014) have both expressed skepticism about the existence of any sort of ‘ought’ that takes into account normative uncertainty. Discussion of their views will have to wait for another time. But we note, first, that even Weatherson acknowledges that his view is *prima facie* ‘extremely unintuitive’. Second, we note that both Weatherson and Harman both argue against this view on the supposition that what we call ‘appropriateness’ is a sort of moral ought. But that is not the established view in the literature: (Lockhart 2000, 24–26), (Ross 2006), (Sepielli 2009, 10), (MacAskill 2014) and (Wedgwood 2013) are all explicit in claiming that what we call ‘appropriateness’ is a type of *rational* ‘ought’. So Weatherson and Harman are at risk of defeating only a straw man.

<sup>12</sup> (Gracely 1996, 301)

<sup>13</sup> (Gustafsson and Torpman 2014).

<sup>14</sup> We will often represent decision-situations using tables. In these tables, actions are listed in the left-hand column, theories and the decision-maker's credences in those theories (represented using percentages) are listed in the top column, and all other cells in the table refer to the choice-worthiness (or normative status) of the action in the cell's row, according to the theory in the cell's column.

<sup>15</sup> (Gustafsson and Torpman 2014, 14).

<sup>16</sup> An additional problem for this account is that it now divorces the theoretical framework for decision-making under normative uncertainty from what we'd think of as the typical situation for real-life decision-makers, who have credence in partial theories or single moral propositions like 'eating meat is wrong', rather than fully-specified moral theories. We thank an anonymous reviewer for raising this point.

<sup>17</sup> (Lockhart 2000, 26) suggests this view, though ultimately rejects it.

<sup>18</sup> There is a standard response to the money-pump argument (E.g. (Schick 1986; Mongin 2000)), as follows. Either Ursula knows what trades are going to happen or she doesn't. If she doesn't, then the objection doesn't have much force: it's unfortunate that she doesn't know what's going to happen in the future; but it's that lack of knowledge that makes her lose money, and there is no reason to doubt Ursula's rationality just because a lack of knowledge makes her lose money. If in contrast she does know what trades are going to happen, then she can foresee what her choices will be at later stages, and use backwards induction to work out what that means her choices should be at the first stage. In the case above, if Ursula knows what trades are going to be made to her, she can *see* that, at Stage 3, she will choose A' (because the choice is between C and A'). Given that she knows that she will trade at Stage 3, then she should refuse to trade at Stage 2 (because her choice, at Stage 2, is effectively between B and A'). But she still prefers B to A (which is her choice at Stage 1). So she will trade at the first step, and then stop. She won't end up strictly worse off than when she started, and so she isn't susceptible to a money pump. However, though this response works in simpler cases, it doesn't work in more complex cases. This has been shown by (Rabinowicz 2000).

<sup>19</sup> (Ross 2006, 743).

<sup>20</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 2006, 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>21</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 choice-worthiness of an option is the average of its choice-worthiness according to the different theories, weighted by the credence in those theories.

<sup>22</sup> (Buchak 2013).

<sup>23</sup> For an argument of this sort, see (Sepielli 2010).

<sup>24</sup> (Weatherson 2014, 148–149).

<sup>25</sup> We thank Amanda Askell for this point, and for the following example.

<sup>26</sup> In fact the contrast of pleasure versus self-realization is a reference to (Hudson 1989, 224) who listed these two quantities 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.

<sup>27</sup> Weatherson hints at this objection in (2002); it is made at length in (Barry and Tomlin 2016). For discussion, see (Sepielli 2010, 103–5).

<sup>28</sup> (Singer 1972).

<sup>29</sup> This is the current best-guess estimate from [www.givewell.com](http://www.givewell.com), a leading charity evaluator (though estimates vary over time).

<sup>30</sup> Weatherson says: "The principle has some rather striking consequences, so striking we might fear for its refutation by a quick *modus tollens*" (2002, 694) 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" (2014, 145). In this, therefore, we agree with him.

<sup>31</sup> We thank an anonymous referee for pushing this objection. See also (Lockhart 2000). Note that what we say will also apply to ‘satisficing’ consequentialist theories: like other theories that posit supererogation, they make a distinction between what is permissible and what is (in some sense) optimal.

<sup>32</sup> We take this classification, and the references below, from (Sepielli 2010 pp. 238–245).

<sup>33</sup> For examples of this type of account, see (Raz 1975) (William 1977) and (Wolf 2009).

<sup>34</sup> For examples of this type of account, see (Portmore 2003), (Zimmerman 1996) and (Chang 2009).

<sup>35</sup> In other, currently unpublished, work, we defend accounts of how to take normative uncertainty into account when those conditions are not met.

<sup>36</sup> This seems to be the view of (Hudson 1989), (Gracely 1996) and (Gustafsson and Torpman 2014).

<sup>37</sup> In other, currently unpublished, work, the first author develops a new account of intertheoretic comparisons.

<sup>38</sup> Our discussion in what follows bears a debt to Ross’s idea of ‘shared parts’ between theories. What a full-fledged account of intertheoretic comparisons would need to do is to explain what *makes it the case* that a certain part of a theory is ‘shared’ between two different theories. Sepielli attempts to cash this view out in one way (Sepielli 2009), but we don’t find his way of doing so plausible. He recants this view in (Sepielli ms), defending an account based on the conceptual role of concepts like ‘ought’ and ‘good’.

<sup>39</sup> This problem is first raised by (Ross 2006, 765).

<sup>40</sup> (Ross 2006, 766).

<sup>41</sup> Though see (Hájek 2003b) for arguments against the standard view.

<sup>42</sup> See (Easwaran 2014) for arguments against using hyperreals in our models of credences. See (Hájek 2003a) for discussion of how invoking infinitesimals fails to help with the ‘fanaticism’ problem within decision theory under empirical uncertainty.

<sup>43</sup> For further discussion of the problems that infinite amounts of value pose for decision-theory, see (Hájek 2003a).

<sup>44</sup> 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. (Beckstead ms) provides discussion.

<sup>45</sup> (MacAskill 2016).

<sup>46</sup> Noting, again, that we are proposing MEC as one part of a more general account of decision-making under moral uncertainty, which would deal with the fact that a decision-making will typically face varying amounts of information from different theories in which she has credence.

## Bibliography

- Barry, Christian, and Patrick Tomlin. “Moral uncertainty and permissibility: evaluating option sets.” *Canadian Journal of Philosophy* 46.6 (2016): 898–923.
- Beckstead, Nick. ms. “Recklessness, Timidity, and Fanaticism.”
- Buchak, Lara. 2013. *Risk Aversion and Rationality*. Oxford University Press.
- Chang, Ruth. “Voluntarist Reasons and the Sources of Normativity”, in David Sobel and Steven Wall (eds.), *Reasons for Action*, Cambridge: Cambridge University Press, 2009, pp. 243–271.
- Easwaran, Kenny. 2014. “Regularity and Hyperreal Credences.” *Philosophical Review* 123 (1): 1–41.
- Gracely, Edward J. 1996. “On the Noncomparability of Judgments Made by Different Ethical Theories.” *Metaphilosophy* 27: 327–32.
- Graham, Peter. 2010. “In Defense of Objectivism about Moral Obligation.” *Ethics* 121 (1): 88–115.
- Guerrero, Alexander A. 2007. “Don’t Know, Don’t Kill: Moral Ignorance, Culpability, and Caution.” *Philosophical Studies* 136: 59–97.
- Gustafsson, Johan, and Tom Torpman. 2014. “In Defence of My Favorite Theory.” *Pacific Philosophical Quarterly*.
- Hájek, Alan. 2003a. “Waging War on Pascal’s Wager.” *The Philosophical Review* 112 (1): 27–56.
- . 2003b. “What Conditional Probability Could Not Be.” *Synthese* 137 (3): 273–323.
- Harman, Elizabeth. 2014. “The Irrelevance of Moral Uncertainty.” In *Oxford Studies in Metaethics*.
- Hudson, James L. 1989. “Subjectivization in Ethics.” *American Philosophical Quarterly* 26: 221–29.

- Jackson, Frank. 1991. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101 (3): 461–82.
- Lockhart, Ted. 2000. *Moral Uncertainty and Its Consequences*. Oxford University Press.
- MacAskill, William. "The Infectiousness of Nihilism." *Ethics* 123.3 (2013): 508–520.
- . 2014. "Normative Uncertainty". DPhil thesis, University of Oxford.
- . "Normative Uncertainty as a Voting Problem." *Mind*, vol. 125, no. 500 (October, 2016), pp. 967–1004.
- . "Smokers, Psychos, and Decision-Theoretic Uncertainty." *The Journal of Philosophy* 113.9 (2017): 425–445.
- MacAskill, William, Krister Bykvist & Toby Ord. Under contract. *Moral Uncertainty*. Oxford: Oxford University Press.
- Mongin, Philippe. 2000. "Does Optimization Imply Rationality?" *Synthese* 124 (1–2): 73–111.
- Oddie, Graham. 1995. "Moral Uncertainty and Human Embryo Experimentation." In *Medicine and Moral Reasoning*, edited by K. W. M. Fulford, Grant Gillett, and Janet Martin Soskice. OUP.
- Portmore, Douglas W. "Position-Relative Consequentialism, Agent-Centered Options, and Supererogation", *Ethics*, vol. 113, no. 2 (January, 2003), pp. 303–332.
- Rabinowicz, Wlodek. 2000. "Money Pump with Foresight." In *Imperceptible Harms and Benefits*, edited by Michael J. Almeida, 123–54. Library of Ethics and Applied Philosophy 8. Springer Netherlands.
- Raz, Joseph. "Permissions and Supererogation", *American Philosophical Quarterly*, vol. 12, no. 2 (April, 1975), pp. 161–168.
- Regan, Donald. 1980. *Utilitarianism and Co-Operation*. Clarendon Press.
- Ross, Jacob. 2006. "Rejecting Ethical Deflationism." *Ethics* 116 (July): 742–68.
- Schick, Frederic. 1986. "Dutch Bookies and Money Pumps." *The Journal of Philosophy* 83 (2): 112–19.
- Sepielli, Andrew. 2009. "What to Do When You Don't Know What To Do." In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, 350. Oxford University Press.
- . "Intertheoretic Comparisons of Value."
- Singer, Peter. 1972. "Famine, Affluence, and Morality." *Philosophy & Public Affairs* 1: 229–43.
- Weatherson, Brian. 2002. "Review of Ted Lockhart's 'Moral Uncertainty and Its Consequences.'" *Mind* 111: 693–96.
- . 2014. "Running Risks Morally." *Philosophical Studies* 167 (1): 141–63.
- Wedgwood, Ralph. 2013. "Akrasia and Uncertainty." *Organon F.* 20 (4).
- Williams, Bernard. "Persons, Character, and Morality", in Amélie Oksenberg Rorty (ed.), *The Identities of Persons*, Los Angeles: University of California Press, 1977, pp. 197–216.
- Wolf, Susan. "Moral Obligations and Social Commands", in Samuel Newlands & Larry M. Jorgensen (eds.), *Metaphysics and the Good: Themes from the Philosophy of Robert Merrihew Adams*, Oxford: Oxford University Press, 2009, pp. 343–367.
- Zimmerman, Michael. *The Concept of Moral Obligation*, Cambridge: Cambridge University Press, 1996.
- . 2006. "Is Moral Obligation Objective or Subjective?" *Utilitas* 18 (04): 329–61.
- . 2008. *Living with Uncertainty: The Moral Significance of Ignorance*. Cambridge University Press.