Moral Uncertainty, Proportionality and Bargaining

Patrick Kaczmarek, Harry R. Lloyd and Michael Plant

Abstract. As well as disagreeing about how much one should donate to charity, moral theories also disagree about *where* one should donate. In light of this disagreement, how should the morally uncertain philanthropist allocate her donations? In many cases, one intuitively attractive option is for the philanthropist to split her donations across all of the charities that are recommended by moral views in which she has positive credence, with each charity's share being proportional to her credence in the moral theories that recommend that particular charity. Despite the fact that something like this approach is already being used by real-world philanthropists to distribute billions of dollars of donations, it is not supported by any of the approaches to moral uncertainty that have been proposed thus-far in the philosophical literature. In this paper, we will develop a bargaining-based approach to moral uncertainty that honors the proportionality intuition in favor of splitting one's donations. We also show how this bargaining-based approach has several further advantages over the best alternative proposals.

§1

Consider:

> *Torn Up*. Jane intends to give away her fortune. Although positive that suboptimal sacrifices are wrong, Jane is torn between two moral views. One view implies that she should fund an initiative providing deworming pills to distant children; the second implies that she should support soup kitchens in her hometown. After many sleepless nights, she is no closer to knowing what the right thing is to do.

Jane's predicament is all too familiar. Each of us has made tough choices while plagued by doubt.

What should Jane do?

When agents are deciding under uncertain conditions, we can distinguish between the 'objective should', the 'subjective should' and the 'super-subjective should'. The 'objective should' describes what should be done given full knowledge of the situation. Suppose the moral view that implies Jane should donate to the deworming initiative is in fact true. If so, then she objectively should fund the deworming initiative, and it would be wrong for Jane to fund local soup kitchens.[1] Doubtless, this is the advice Jane prizes most and wishes she had to follow in *Torn*

---

[1] However, even if impermissible, donating to soup kitchens could still be praiseworthy, or at least not blameworthy (Pummer 2021).

Up; however, the moral facts are opaque to her. By contrast, the two remaining senses of 'should' guide deliberation by virtue of being sensitive to an agent's false and gappy beliefs (Muñoz and Spencer 2021, 77). Following Hedden (2016) and Sung (2022), we understand the 'subjective should' as being sensitive to an agent's descriptive uncertainty, whereas the 'super-subjective should' is sensitive to her descriptive *and moral* uncertainty.[2] This last sense of 'should' is most relevant to Jane's situation. After all, try as she did, Jane could not glean the objectively right thing to do. All Jane has to go on are her patchy beliefs about what is good or bad, permissible or wrongful, supererogatory etc. Unless otherwise stated, by 'should' we will mean 'super-subjectively should' from this point onward.

Our own intuition is that Jane should split her donations in *Torn Up*. She should give some portion of her donations to the deworming initiative and the rest to soup kitchens in her hometown, where the distribution of her donations corresponds to her credences in the two moral views that Jane is torn between. Let's call this response 'Proportionality'. Stated more generally, Proportionality is the view that if some decision maker has x% credence in some moral theory, then she should use x% of her resources in the manner recommended by that particular moral theory.[3]

Many people, we gather, feel the same about *Torn Up* and cases like it; so much so that Proportionality might be said to be part of common sense.[4] But even if the intuition is far from ubiquitous, it is in fact regularly relied upon to decide the fates of millions of people, many of whom are poor, vulnerable to disease and children. In particular, some effective altruists seem to rally behind this intuition.[5] This social movement counts Dustin Moskovitz and Cari Tuna among its ranks, who have promised to give away billions of dollars in (apparent) accordance with Proportionality.[6]

And yet, despite the practical importance of doing so, the widespread practice of diversifying donations when morally uncertain has gone unexamined by philosophers; even by

---

[2] Some philosophers deny there are subjective norms to guide an agent's deliberations under conditions of moral uncertainty (Harman 2014; Weatherson 2019). Important though these arguments are, this paper is not the right place to engage with them. We focus on whether these subjective norms, if there are any, could accommodate intuition about *Torn Up*.

[3] By 'recommended' we simply mean that the moral theory describes the option as a permissible object of choice. Some theories will deem more than one option permissible while smiling most on one of these (i.e., a supererogatory option). As we will discuss in §4.e, Proportionality grants the uncertain decision maker leeway here, which we take to be an attractive feature of the view. Thanks to [name removed for blind review] for pressing us to clarify this point.

[4] [anecdotal evidence removed for blind review]

[5] Effective altruism community members search for ways to do the most good and then put them into practice; see https://www.effectivealtruism.org/articles/introduction-to-effective-altruism. Arguably, its moral foundation is the (Weak) Avoid Gratuitous Worseness Principle, which states "It is wrong to perform an act that is *much worse* than another, if it is *no* costlier to you to perform the better act, and if all other things are equal" (Pummer 2016, 84; cf. Sinclair 2018). See also (Pummer 2023).

[6] Open Philanthropy, the grant-making organization charged with executing the philanthropic plans of Moskovitz and Tuna, subscribes to an approach they label 'worldview diversification', described here: https://www.openphilanthropy.org/research/worldview-diversification/. The rationale given is somewhat imprecise, but the relevant features are that they split their credence across different 'worldviews'—which reflect various positions one can take on both moral and empirical uncertainties—and then allocate *some* resources to each of those worldviews.

those who would consider themselves card-carrying effective altruists.[7] The paper rectifies this oversight. We defend a novel account for handling moral uncertainty honoring the intuition behind Proportionality.[8]

Here is the plan. In §2, we will present two challenges for constructing this account. First, Proportionality does not cover cases where the agent faces a choice between discrete options, as opposed to a resource distribution case like *Torn Up*. Second, there is a class of cases—so-called 'Jackson cases'—where Proportionality could be applied but delivers counterintuitive verdicts.[9] Jackson cases drove many of our colleagues into the arms of Maximize Expected Choice-Worthiness, a rival account of appropriate behavior under conditions of moral uncertainty. We will put pressure on their argument in §3. Finally, we will develop and defend a bargaining approach across §§4-6.

<h1 style="text-align:center">§2</h1>

Proportionality strikes many as a plausible solution to the problem of "where to give" when morally uncertain. But we face moral uncertainty in non-donation cases too, regarding the wrongness of eating meat, breaking promises, diverting trolleys into innocent bystanders and so on.

As this section demonstrates, Proportionality tends to offer bad advice on these other quandaries. But if we must appeal to one or more additional subjective norms for reasonable guidance in these kinds of cases, then this brews doubts about Proportionality. Is there a unified account that explains all possible cases?

2.a. Proportionality is incomplete

The most immediate challenge to Proportionality is that it has no guidance to offer in a variety of choice situations. In general, Proportionality does not cover those cases where the agent faces a choice between discrete options, as opposed to a resource distribution case like *Torn Up*. Consider:

> *Old-Timey Villain.* A dastardly villain has placed two strangers, bound and gagged, along the tracks and fled. A runaway trolley is barrelling down the track, and it will soon kill them both if you do nothing. Standing beside you is a man, George, wearing a heavy backpack. If you push him into the path of the trolley, their combined weight will cause

---

the trolley to come to a complete stop before killing the pair of strangers, but George will unfortunately be crushed to death.[10]

Suppose that you are torn between two moral theories. You have 10% credence in a moral theory that prescribes shoving George into the trolley path, since doing so is best. But, you also have 90% credence in a moral theory that proscribes doing so (because George has not waived his right against bodily harm).

What does Proportionality recommend?

Nothing; this is because it is impossible to commit to both pushing George *and* not pushing him into the path of the trolley. Indeed, even if you had the option of dangling George's legs over the tracks, such that he survives and the trolley stops before killing the second but not the first stranger, doing so neither *partially* violates his rights nor treats *only part of George* as a mere means. Rather, doing so violates his right and fails to treat him as an end in himself. And so, even when armed with this expanded choice-set you will still be unable to partially satisfy both moral views in *Old-Timey Villain*.

Many of the moral situations that ordinary people will face in life involve choosing between discrete options. Proportionality either goes silent or asks the impossible of us in these discrete choice scenarios. This is highly troublesome.

To our minds, the decision maker should *not* push George onto the tracks in *Old-Timey Villain*. We will revisit the question of how to square this result with Proportionality in §5.

2.b. Proportionality is reckless

Perhaps there's some relief to be found in telling ourselves that the first challenge does not yet reveal a fatal error in choosing to split donations proportionally when morally uncertain. However, there is a second, no less severe, problem awaiting those who were unshaken by the first.

We begin by looking at a classic puzzle, *Miners*, and pulling out the main lesson that it teaches.

> *Miners.* There was a disaster in the quarry, and 100 miners are trapped in Shaft A; the nearby Shaft B is empty. You know that, if you do nothing, the shafts will partly flood and 10 miners will die. You also know that, if you block the shaft where the miners are, you will save all 100. And if you block the empty shaft, the other will totally flood, drowning all 100. But your evidence doesn't tell you where the miners are; for you, it's a 50/50 guess (Muñoz and Spencer 2021, 78).

What should you do?

It isn't the case that you objectively should refrain from blocking either shaft. After all, you know that if the miners are trapped in Shaft A, then you objectively should block Shaft A. And you also know that if the miners are trapped in Shaft B, then you objectively should block

---

[10] We took pity on the so-called "fat man", a fabled victim of trolleyology, and adapted a case from (Mogensen 2016, 215-216).

Shaft B. Wherever these miners are located, blocking neither mine shaft is sure to be the wrong thing to do.

Yet, your doxastic attitudes being what they are, it is *reckless* to block either mine shaft; because you don't know their location, you are as likely to kill the miners as you are to rescue them. The lesson we are meant to learn in *Miners* is this: "you (subjectively) shouldn't even try to do as you objectively ought, because you don't know which shaft you objectively ought to block—and a wrong guess spells disaster" (Muñoz and Spencer 2021, 79).

Miners is a Jackson case. The following criteria make for a Jackson case: (a) the agent should choose an option that is suboptimal; (b) the agent knows that the option she should choose is suboptimal; and (c) it would be unacceptably reckless for the agent to choose any other option (Field 2019, 394). The purpose of having gone through the *Miners* exercise was to establish as much.

Notice, Proportionality *never* recommends putting *any* resources towards outcomes known to be objectively wrong (or rather, actions which every moral theory that you have positive credence in deems impermissible). So, Proportionality will never honour the lesson from *Miners*. As such, Proportionality is objectionably insensitive to the stakes described by moral theories. Consider:

> *Mining Safari.* You know all of the following. There was a disaster in the quarry: 10 giraffes are trapped in Shaft A and 20 canaries are trapped in Shaft B. There isn't enough time to fully block both shafts. If you block Shaft A, the giraffes will be saved, but the other shaft will totally flood, killing the canaries. If you block Shaft B, the canaries will be saved, but the other shaft will totally flood, killing the giraffes. You could partially block each shaft, but bricks and other deadly debris will then get washed into both mine shafts by the water, making the flood that much more deadly and killing everyone inside. If you block neither shaft, 6 giraffes and 12 canaries will survive.

Suppose that you are equally torn between Peter Singer's utilitarianism, according to which all creatures with moral standing share the same moral status (Singer 2009), and Shelly Kagan's hierarchical approach, which assigns a lower moral status to canaries than giraffes (Kagan 2019). For concreteness, let's assume that a canary's claim has 1/4 the strength of a giraffe's claim.

Table 1 describes how Singer and Kagan's views describe the outcomes in *Mining Safari*.

Block:	Singer's View	Kagan's View
A	10	10
B	20	5
Neither	16	8
Both (partially)	0	0

Table 1. The upside in Mining Safari.

Singer's view recommends blocking Shaft B; meanwhile, Kagan's view recommends blocking Shaft A.[11]

What should you do?

Proportionality impels you to (partially) block both shafts, which is absurd, since all the animals die. More so, although they disagree about the ranking order of the alternatives in *Mining Safari*, Singer and Kagan's views each, internally, recognize this as the worst possible outcome in the situation. This feature of the case seems relevant, and perhaps that's where Proportionality goes wrong.

Suppose that it was instead Peter and Shelly, rather than you, who stumbled into *Mining Safari*. What would they do? We cannot imagine that either of them would dig their heels in, refusing to alter course in light of the other's preference. Peter and Shelly would recognize, in other words, that they cannot singlehandedly determine the outcome. This brings out the underlying flaw in Proportionality: *it overlooks how the various recommendations from each theory combine to bring about some final outcome*.

We can begin to patch this flaw by proposing that your decision procedure should reflect what flesh-and-blood Peter and Shelly would actually do. Their theories should be made to, in a sense to be explained, *cooperate*. Let's suppose that each view you are torn between is assigned a representative in your practical deliberations, and that each representative tailors their recommendation to account for the preferences of other representatives. In this case, the first representative champions Singer's view while the second champions Kagan's view. What would your Kagan representative recommend in light of what the Singer representative prefers happen with her share of your resources? He would not recommend blocking Shaft A conditional on the Singer representative blocking Shaft B, since doing so guarantees the worst possible outcome. What would your Singer representative recommend in light of what the Kagan representative prefers happen with his share of your resources? Similarly, she would not recommend blocking Shaft B conditional on your Kagan representative blocking Shaft A, since doing so guarantees the worst possible outcome. Thus, we can rule out partially blocking both mineshafts in response to *Mining Safari*. However, we have not yet fleshed out this basic idea in enough detail to determine which of the remaining options these representatives would actually recommend (that detail will come in §4 below).

Although we seem to be on the right track by viewing the problem of what to do when morally uncertain as a cooperation problem, the idea just sketched falls short of a satisfying solution. This is because failures of cooperation can crop up even when representatives are made aware of one another's preferences. The problem can be illustrated with the aid of another toy example.

> *Torn Up**. Jane intends to give away her life savings. If she funds a fertility initiative, two additional children will be born. Alternatively, Jane can fund a contraception initiative that results in two fewer pregnancies. She is equally torn between impersonal total

---

[11] In fact, these verdicts depend on several additional features that we brushed aside for simplicity's sake, such as how good an average giraffe's life is for them, the lifespan of a canary, their degrees of psychological capacity (at least, on Kagan's view) and so forth.

utilitarianism and anti-natalism.[12] The former implies that Jane has strong all-things-considered reason to fund the fertility initiative, and equally strong all-things-considered reason not to fund the contraception initiative. By contrast, anti-natalism implies that Jane has strong all-things-considered reason to fund the contraception initiative, and equally strong all-things-considered reason not to fund the fertility initiative. Both of these moral theories agree that Jane has *some* all-things-considered reason to fund the Against Malaria Foundation, which supplies insecticide-treated bed nets to children at risk of contracting, and dying from, malaria. However, both views also agree that funding the Against Malaria Foundation would be wrong *qua* suboptimal. After many sleepless nights, Jane is no closer to knowing what the right thing is to do.

If Jane gives the fertility and contraception initiatives an equal share of her money, they will balance each other out, leaving the world exactly as she found it (as far as these moral theories are concerned). Given this, splitting her donations is no more valuable than doing nothing, frittering away her fortune.

As above, suppose that Jane deliberates as if she had two representatives tailoring their recommendations in *Torn Up*.* Whatever the anti-natalist representative does, her total utilitarian representative prefers to fund the fertility initiative. After all, if the anti-natalist representative were successful in preventing the birth of, say, Elroy, the world would be worse on balance, according to the total utilitarian representative, given there would be less happiness in the population. From the total utilitarian representative's point of view, maintaining the status quo by creating, say, Judy is more important than bed nets. And if the anti-natalist representative gives out bed nets, then the total utilitarian representative still prefers creating Judy over supplying bed nets. Jane's anti-natalist representative similarly prefers funding the contraception initiative whatever the total utilitarian representative chooses to do. So, together they squander Jane's donations. And yet, both the total utilitarian representative and the anti-natalist representative agree that bed nets are better than nothing at all.

Parfit called this an 'Each-We Dilemma' (1984, 91).[13] If each of Jane's theory representatives produces the best outcome they can individually, they produce a worse outcome collectively.

---

[12] Total utilitarianism states that one should bring about the outcome in which there would be the greatest quantity of happiness (Parfit 1984, 387). On this view, we should create happy people, since doing so increases the total sum of happiness. (Some find this conclusion disturbing; they believe how good a state of affairs is depends only on how good it is *for* persons (Roberts 2003, 1). As Jonathan Bennett bemoaned, "As well as deploring the situation where a person lacks happiness, [total utilitarians] also deplore the situation where some happiness lacks a person" (Bennett 1978, 63-64).) Anti-natalism holds the polar opposite; it is gravely wrong to create people, even if they are on balance happy, since the anti-natalists argue that we harm people terribly when we create them but do not benefit them at all (Benatar 2006; cf. Pallies forthcoming). Our grasp of Benatar's view was greatly improved by Elizabeth Harman's review of it (2009).

[13] *Torn Up** is a *pure* Each-We Dilemma since it can be solved by improving coordination between agents. See (Temkin 2022, 238-249) for another type of Each-We Dilemma, one that cannot be solved in this fashion (Clark and Pummer 2019, 30).

We believe that Jane should donate all of her savings to the Against Malaria Foundation in *Torn Up**. And we don't seem to be alone in thinking this; Toby Ord (2015) defends the same conclusion in scenarios where *distinct* people rather than imaginary representatives hold different moral views and coordinating would be mutually beneficial. He refers to this as 'moral trade'. This suggests that a promising approach to handling moral uncertainty may be to treat it as a case of intra-personal bargaining, where we imagine what the representatives of the different moral theories that one believes in would do, if given the opportunity to coordinate their actions.

2.c. Recap

§2 surveyed the main challenges for constructing a comprehensive account for handling moral uncertainty that honors the practice of diversifying one's donations when torn between competing moral theories. Viewing the problem of what one ought to do when morally uncertain as a cooperation problem between moral theories seems promising. At first blush, a sophisticated bargaining approach to moral uncertainty seems well-placed to explain all of the cases in this paper.

We will further motivate the idea in §4. First, however, we introduce Maximize Expected Choice-Worthiness—a popular rival to the approach that we will defend in this paper.

§3

The most popular decision procedure in the literature on moral uncertainty is Maximize Expected Choice-Worthiness (hereafter 'MEC'), where the 'choice-worthiness' of some action according to a moral theory is the strength of the decision maker's all-things-considered reasons in favor of performing that action according to that moral theory (MacAskill and Ord 2020, 329). The 'expected choice-worthiness' of some action is a weighted average of its choice-worthiness according to each of the theories in which the decision maker has credence, where each theory's weight in the average is the decision maker's credence in that theory.

Thus, MEC says that we should handle moral uncertainty in the same way as expected utility theory says that we should handle empirical uncertainty. In fact, several advocates of MEC regard this analogy with standard decision theory as a reason to endorse MEC. For instance, MacAskill, Bykvist and Ord claim that since "expected utility theory is the standard way that we should handle empirical uncertainty … maximizing expected choiceworthiness should be the standard account of how to handle moral uncertainty" (2020, 47-48). In a similar vein, Christian Tarsney maintains that treating moral and empirical uncertainty "differently when we are not forced to is at least prima facie inelegant and undermotivated" (2021, 172; cf. Sepielli 2010, 75-78).

Unfortunately, however, we think that there are some disanalogies between moral and empirical uncertainty, which call into question the argument from analogy in favour of MEC. To be clear: we do not think that these disanalogies constitute a fatal blow to MEC. All we hope to show in this section is that the case isn't open and shut. That's enough for our purposes; all we want to show is that proposing an alternative decision procedure for handling moral uncertainty isn't a non-starter.

Perhaps the most important disanalogy between empirical and moral uncertainty concerns intertheoretic choice-worthiness comparisons. In paradigm cases of decision making under empirical uncertainty, the goodness or badness of each of the various possible outcomes can be measured on some shared evaluative scale. For instance, the value of the different possible outcomes at a casino table can be measured in terms of dollars won or lost. And the value of several different possible plays in gridiron football can be measured in terms of net points won or lost. In the absence of this kind of comparability, it would simply be impossible to calculate the expected value of any particular action.

Unfortunately for MEC, it remains deeply controversial whether it is possible to make intertheoretic choice-worthiness comparisons across several different moral theories. Advocates of MEC such as MacAskill, Bykvist and Ord have argued that the choice-worthiness scores assigned to options by different moral theories might all be cardinally measurable on a shared "universal scale" of choice-worthiness (2020, chapter 5). On the other hand, critics of intertheoretic comparisons have argued that "it is part of the very nature of a moral system that it presents a way of viewing reality, and that the differing visions of different systems cannot be directly compared" (Gracely 1996, 328). For instance, imagine trying to compare absolutist deontology against scalar utilitarianism. These two moral theories don't even use the same deontic categories: absolutist deontology sees the world only in terms of permissions and prohibitions, whereas scalar utilitarianism sees the world only in terms of betterness and worseness of outcomes in terms of aggregate utility. It strikes many working in the field as implausible to suppose these two moral theories both rank actions on a shared 'universal scale' of choice-worthiness.[14] This is an important disanalogy between empirical and moral uncertainty.[15]

A second problem with the argument from analogy is that selecting a decision procedure that is designed to recommend optimal *gambles* strikes us as *prima facie* much less appealing in the moral uncertainty case than it is in the descriptive uncertainty case. Instead, we think it is more attractive to adopt an approach to moral uncertainty that is designed to select optimal *compromises* between the moral theory in which one has positive credence. MacAskill himself suggests an alternative analogy between moral uncertainty and social choice, and explicitly emphasizes the idea of compromising:

> The formal structure of the two problems is very similar. But the two problems are similar on a more intuitive level as well. The problem of social choice is to find the best compromise in a situation where there are many people with competing preferences. The problem of [moral] uncertainty is to find the best compromise in a situation where there are many possible normative theories with competing recommendations about what to do (2016, 977).

---

[14] Other critics of intertheoretic choice-worthiness comparisons include (Broome 2012; Gustafsson and Torpman 2014; Hedden 2016; Gustafsson 2022; [removed for blind review]).

[15] Another disanalogy between empirical and moral uncertainty is that how one should handle empirical uncertainty is one of the matters about which one can be morally uncertain—but not *vice versa*. For a discussion of the importance this disanalogy see [redacted for blind review].

We ourselves develop this alternative analogy in §4 of this paper.

A final problem with the argument from analogy is that it torpedoes Proportionality. According to MEC, it is rarely, if ever, correct to split one's donations according to Proportionality. In general, MEC only ever implies the permissibility of proportionally diversifying donations as a matter of coincidence, such as in cases where all of the available options in some choice situations are maximally choice-worthy in expectation, or in certain very particular cases where the returns to donating to every theory's favored charities diminish at exactly the right rate [removed for blind review].

In this section, we have argued that the case in favour of MEC is far from open and shut. The argument from analogy in favour of MEC is in tension with Proportionality, is less *prima facie* plausible than MEC's advocates assert and papers over a disanalogy between moral and empirical uncertainty concerning the plausibility of value comparisons. In the next section of this paper, we motivate an alternative *bargaining* approach to handling decision making under moral uncertainty.

§4

An observation, almost too obvious to make, is that sometimes we agree with each other about what to do, and sometimes we don't. Where we agree, little more needs to be said: we act. Where we disagree, we often bargain to see if we can find an acceptable compromise. Indeed, bargaining is a skill we develop almost before we can walk. We begin with requests to our parents: requests that they'll often grant in exchange for us doing what they want—such as eating our greens and sitting quietly in church. Negotiation continues apace thereafter. We learn how to compromise with sweethearts, friends and enemies. Industry titans and politicians alike wheel and deal. Striking a bargain is fundamental to our lives as social beings, something so familiar we sometimes barely recognize that we are doing it. Sometimes, bargaining does not work, or is impractical. In some of these contexts, we may turn to voting and imposing decisions on others they do not want. *In extremis*, we resort to force.

Given the ubiquity of bargaining as a means of resolving disagreements, it is perhaps surprising that, whilst voting theory is widely referenced in the moral uncertainty literature, bargaining is not. In the paper comparing moral uncertainty to social choice that we quoted from in §3 above, MacAskill adopts a voting-theoretic approach. Similarly, an early, but still-underdeveloped proposal for moral uncertainty was the moral parliament, where theory representatives vote on what decisions to take.[16] Note, however, that bargaining may equally fit the bill for finding a compromise between normative theories. We've already hinted at its potential in §2 of this paper.

The aim of this section is to develop a bargaining approach to moral uncertainty. Our framework takes its inspiration not from parliaments, but from the marketplace; we call it the Moral Marketplace theory ('MMT'). MMT mimics the kind of interactions that take place in free markets with well-defined property rights over goods and labor. According to MMT, how a morally uncertain decision maker should act in any given choice situation is determined by a certain economic model of that choice situation. Each moral theory is modelled as an economic

---

[16] The most recent treatment of the moral parliament approach can be found in Newberry and Ord's (2021).

agent, who is endowed with a share of the decision maker's resources proportional to the decision maker's credence in the corresponding moral theory. This representative of the theory is modelled as the owner of these resources and can use those resources however they see fit. Each representative's preference ranking over how resources are used overall is identical to the choice-worthiness ranking of the moral theory that it represents. To this end, we will restrict our attention to cases where all of the moral theories in which our uncertain decision makers have credence exhibit a certain kind of cardinal structure. More specifically, we only consider theories that can be presented by *interval-scale* choice-worthiness functions.[17]

Representatives can make deals with each other, but they don't have to. Because they have the right to their own resources, a pair of representatives will make a deal if and only if that deal would be mutually beneficial—that is, if both believe the bargain is better for them than acting unilaterally. An important issue, one we will come back to, is determining in different contexts the appropriate 'disagreement point': what happens if the representatives cannot agree.

Here's how we proceed. We start by explaining what MMT recommends in cases where resources are divisible, for instance allocating resources to charity. Along the way, we highlight how MMT has several attractive features: it vindicates proportionality in certain cases (§4.a), delivers the desired verdict in *Torn Up** (§4.b), does not require intertheoretic comparison of choice-worthiness (§4.c) and avoids the challenges of both fanaticism (§4.d) and demandingness (§4.e). From there, we move on to consider cases where resources are non-divisible (§5) before addressing the main challenge facing proponents of bargaining approaches to handling moral uncertainty: the *problem of small worlds* (§6). We then conclude.

4.a. Divisible resources, no bargains available
For a simple illustration of these ideas, consider the *Torn Up* thought experiment from §1 of this paper. In *Torn Up*, Jane is torn between two moral theories, one of which recommends donating to a deworming initiative, and the other of which recommends donating to soup kitchens in her hometown.

MMT suggests that Jane should model these two moral theories as two economic agents, each of whom is initially endowed with a share of Jane's fortune proportional to Jane's credence in the corresponding moral theory. If these two representatives wished to, they could make contracts with each other. And they can also choose to spend their endowments in any of the

---

[17] A unit increase on an interval represents a certain fixed amount of the underlying thing being measured, regardless of where on that scale that unit increase occurs. For instance, °F and °C are both interval-scale measures of temperature because an increase of 1°F or 1°C represents the same amount of extra heat at freezing point as it does at boiling point.

We restrict our attention to interval-scale measurable moral theories only for sake of simplicity. Restricting one's attention to a certain subclass of moral theories is a common strategy for making progress in the existing literature on moral uncertainty. For instance, MEC is only applicable to uncertainty over theories that can all be represented by interval-scale and intertheoretically comparable choice-worthiness rankings. Advocates of MEC have gone on to propose alternative decision procedures for cases where these two conditions are not satisfied, such as MacAskill's *Borda count* voting-theoretic approach (MacAskill 2016). Similarly, Tarsney (2018) restricts his attention to cases of moral uncertainty over different versions of absolutist deontology when proposing a *stochastic dominance* approach. See also (Greaves and Ord 2017).

ways open to Jane. As it happens, in *Torn Up* these two representatives do not have anything to gain by making contracts with each other. The first representative just wants to donate all of her endowment to the deworming initiative, and the second representative just wants to donate all of her endowment to local soup kitchens. Hence, according to MMT, Jane should split her fortune proportionally between deworming pills and soup kitchens.

As this discussion of *Torn Up* makes plain, MMT 'builds in' Proportionality as, in some sense, the 'default response' to cases of moral uncertainty in which the decision maker is deciding how to distribute some continuously divisible resource. MMT deviates from Proportionality only in cases where some alternative resource allocation is a Pareto improvement over the proportional one (we will discuss one such case in §4.b below). Moreover, MMT supplies us with a principled reason for this result. According to MMT, each theory's representative is initially entitled to a proportional share of the decision maker's resources. Thus, each representative always has the option to spend its share of these resources in the manner recommended by the theory that it represents. Each representative will only agree to a contract if it represents an improvement over this proportional response.

4.b. Divisible resources, a successful contract

In some cases, however, MMT will deviate from Proportionality. For instance, consider the *Torn Up** case introduced in §2.b. In that case, if Jane's theory representatives each spent their endowment on the initiative that they regard as optimal, then the total utilitarian representative would donate her endowment to the fertility initiative, and the anti-natalist representative would donate her endowment to the contraception initiative. Each representative regards this overall use of Jane's resources as no better than Jane doing nothing at all.

By contrast, consider a possible outcome in which Jane's total utilitarian and anti-natalist representatives both agree to donate their endowments to the Against Malaria Foundation. Each of these two representatives regards this outcome as better than doing nothing. Hence, each representative regards this outcome as better than the outcome in which each representative spends their endowment on the initiative that they regard as optimal. It is in each representative's interests to enter into a contract with the other representatives which stipulates they will both donate their endowments to the Against Malaria Foundation.

In cases like this, where an agent's theory representatives stand to gain from forming contracts with each other, MMT will need to provide a precise bargaining 'solution concept' to tell us which contract these representatives will agree to. Perhaps the most well-known such solution concept—and the one that we will adopt in this paper—is the *Nash bargaining solution*.

The Nash bargaining solution is the bargaining solution that uniquely satisfies Nash's (1950) four plausible axioms on the outcomes of good-faith (referred to as 'cooperative') bargaining procedures:[18]

1. *Scale invariance*: any positive linear rescaling of any bargainers' utility functions should not alter the bargaining solution.

---

[18] The Nash bargaining solution can also be justified as the limiting outcome of several diachronic, 'non-cooperative' models of the bargaining process (Binmore, Rubinstein and Wolinsky 1986).

2. *Pareto optimality*: no feasible alternatives should Pareto dominate the bargaining solution.
3. *Symmetry*: if every bargainer has the same utility function and disagreement utility, then every bargainer should have the same utility in the bargaining solution.
4. *Independence of Irrelevant Alternatives*: eliminating an element from the set of feasible outcomes should only make a difference to the bargaining solution if the eliminated outcome would itself have been selected as the bargaining solution had it not been eliminated.

In a case like *Torn Up** with only two representatives, an act A is a Nash bargaining solution *iff* setting a = A maximizes

$$\big(CW_1(a) - CW_1(d)\big)\big(CW_2(a) - CW_2(d)\big)$$

where CW_1 is an interval-scale choice-worthiness function for the first moral theory, CW_2 is an interval-scale choice-worthiness function for the second moral theory and d (i.e., the 'disagreement point') is what will happen if the representatives cannot agree to a contract.[19] In other words, d is the proportional outcome in which each representative uses its endowment in the manner recommended by the theory that it represents.

One attractive feature of the Nash bargaining solution is that (all else being equal) it favors equal divisions of the choice-worthiness gains to be had from trade between theory representatives. For example, suppose that two bargainers are choosing between an option A that gives each bargainer a utility gain of 4 over the disagreement point and another option B that gives the bargainers utility gains over the disagreement point of 2 and 6 respectively. Under option A, the value of the Nash maximand is 4 × 4 = 16, whereas under option B, the value of the Nash maximand is 2 × 6 = 12. Hence, as desired, the Nash bargaining approach prefers option A over option B.

In the case of *Torn Up**, CW_1 will be total utilitarianism's choice-worthiness function, CW_2 will be anti-natalism's choice-worthiness function and d will be outcome in which half of Jane's fortune is donated to the fertility initiative, and half of Jane's fortune is donated to the contraception initiative. Let f, c and m denote the proportions of Jane's fortune that she will donate to the fertility, contraception and malaria initiatives respectively.

Since Jane's fortune is small relative to global spending on fertility, contraception and malaria, it is reasonable to assume that if Jane increases her spending on one of those charities by some factor k, this will increase Jane's impact in promoting the goals of that charity by the same factor. Thus, we can assume that CW_1 and CW_2 are *linear* in f, c and m. Together with our original specifications in *Torn Up**, this suggests something like the following specifications for CW_1 and CW_2:

$$CW_1 = 10f - 10c + 3m$$
$$CW_2 = -10f + 10c + 3m$$

---

[19] Note that in order to multiply the choice-worthiness values of two moral theories, we need not assume that these theories are intertheoretically unit-comparable. For a demonstration of this feature of MMT, cf. §4.d below.

The disagreement point d in *Torn Up** corresponds to $\langle f = 0.5, c = 0.5, m = 0 \rangle$. Hence, $CW_1(d)$ = $CW_2(d)$ = 0. Thus, the Nash bargaining solution is the choice of f, c and m that maximizes

$$CW_1 \times CW_2 = (10f - 10c + 3m) \times (-10f + 10c + 3m)$$

(1)

As desired, this solution is $\langle f = 0, c = 0, m = 1 \rangle$. MMT recommends that Jane should donate all of her fortune to the malaria initiative.

4.c. MMT does not depend on intertheoretic comparisons
Having laid out the mechanics and illustrated the functioning of MMT in a couple of cases, we now mention several of its theoretical advantages, starting with the fact it does not require intertheoretic unit comparisons.

As we pointed out in §3.a above, MEC requires us to be able to make intertheoretic choice-worthiness comparisons. In order to decide whether, say, a 10% chance of acting impermissibly according to absolutist deontology is a price worth paying for a 90% chance of acting optimally according to utilitarianism, one needs to be able to commensurate between the choice-worthiness values at stake in this decision according to absolutist deontology and utilitarianism.

By contrast, however, the bargaining approach does not require these kinds of intertheoretic comparisons. Two agents can bargain with each other without having to first establish some kind of exchange rate between their utility functions. All that is required for optimal bargaining is knowing every bargainer's preference structure over potential agreements. Trying to compare different bargainers' levels of satisfaction on some shared 'universal scale' is irrelevant to the bargaining process.

Insofar as intertheoretic choice-worthiness comparisons seem dubious (§3.a above), this is an important advantage of the bargaining approach.

4.d. MMT resists fanaticism
An agent is said to be fanatical if he judges a lottery with tiny probability of arbitrarily high value as better than the certainty of some modest value (Wilkinson 2022, 447).[20] Some approaches to handling moral uncertainty are fanatical about choice-worthiness, including MEC.

To illustrate, consider:

> *Lives or Souls.* You are supremely confident you should give to the Against Malaria Foundation, where your donation will save one child's life. But you have seen evidence that the Against Hell Foundation reliably converts one person to a certain religion, purportedly saving their soul from eternal damnation. You have almost no credence in that

---

[20] 'Fanaticism' was coined by (Bostrom 2011). It has since also been referred to as 'recklessness' (Beckstead and Thomas, 2023).

religion, but you accept that saving a soul is, on that religion, astronomically more valuable than saving a life.

Because the stakes are so much higher on the religious view, even a small credence in that religion's truth threatens to take hostage your decision making under conditions of uncertainty. This is irksome.

By contrast, the bargaining approach to moral uncertainty has principled grounds for avoiding fanaticism (Greaves and Cotton-Barratt 2023, §8). A model that represents the moral theories in which the decision maker has credence as agents bargaining with each other is unlikely to recommend as appropriate an option that one low-credence theory regards as highly choice-worthy, but that every other theory regards as not-at-all choice-worthy. Instead, the bargaining approach is much more likely to recommend an option that every positive-credence theory regards as moderately choice-worthy (if an option like this is available). The outcome of some bargaining process must be an option that is unanimously acceptable to all of the bargainers.

Indeed, it is easy to illustrate mathematically that the Nash bargaining approach described above is not fanatical with respect to choice-worthiness (cf. §4.b). For instance, imagine that CW_1 is the same as before (i.e., $10f - 10c + 3m$), but CW_2 is now scaled up by a factor of ten to $-100f + 100c + 30m$. Then the Nash bargaining solution in *Torn Up** will be the choice of f, c and m that maximizes

$$(10f - 10c + 3m) \times (-100f + 100c + 30m)$$

(2)

However, this expression can be rewritten as

$$10 \times (10f - 10c + 3m) \times (-10f + 10c + 3m)$$

(3)

which is simply ten times expression (1).

Thus, any choice of f, c and m maximizes expression (2) *iff* it maximizes expression (1). Multiplying CW_2 or CW_1 by any positive number does not alter the Nash bargaining solution.[21]

The dust has yet to settle on whether fanaticism is wrongheaded or right but tough to swallow.[22] We take it to be a virtue of the bargaining approach that it does not imply fanaticism.

4.e. MMT preserves moral latitude

Theories like MEC that underwrite 'dominance arguments' can be highly restrictive on morally uncertain agents. To illustrate, consider:

> *New Nets or Wheels?* George is not a rich man and receives a letter from the Against Malaria Foundation, which describes (truthfully and

---

[21] This result also follows directly from the scale invariance axiom.

[22] See (MacAskill, Bykvist and Ord 2020, 150ff) for a defense of adopting the latter stance.

in excruciating detail) how he could save a child's life by donating the few thousand dollars that he has squirreled away for his dream car to their cause instead.

What should he do?

Suppose that George is torn between two moral theories. He is confident of the truth of some commonsense moral theory, which tells him that, although he is not required to send the money to the Against Malaria Foundation, he is permitted to venture beyond the call of duty. But George isn't totally sold; he is somewhat sympathetic to Singer's brand of utilitarianism, according to which you should give away most of your wealth to desperately needy strangers (Singer 1972; Unger 1996). He can't shake the inkling that it's seriously wrong to fail to save a child's life; that continuing to drive a beat-up Honda a little longer wouldn't be the end of the world.

But notice, the risk of wrongdoing in *New Nets or Wheels?* is lopsided. Failing to send the money might be gravely wrong, whereas sending the money is sure to be right (that is, doing so is permissible on both moral views in which George finds purchase). By his own lights, there is no chance of doing something gravely wrong by donating to the Against Malaria Foundation. Since donating dominates buying his dream car, George seems pressed to send all the money. This holds no matter how little stock he puts in that inkling, provided that George assigns it some non-trivial weight.

This is intuitively upsetting. It seems that accounting for moral uncertainty will lead us straight to the Singerian conclusion we should donate all our (spare) resources to charity.[23]

What does MMT have to say about this case?

To begin, notice that *New Nets or Wheels?* is unlike the previous cases. The new feature is that one of the representatives is *indifferent* to making a contract with his fellow representatives. We can imagine the Singer representative jumping up and down, imploring the representative of the commonsense moral view to donate his endowment to the Against Malaria Foundation, whilst the commonsense representative looks back, arms folded and non-plussed. The Singer representative doesn't have anything with which to win over the commonsense representative. Although he would be happy to enter into an agreement like this, such a contract doesn't improve his lot over the disagreement point. He would be equally happy to spend his endowment on a new car.

In cases like this, it is natural to stipulate that either of these two possible uses of the commonsense representative's endowment would count as appropriate by the lights of MMT. In other words, on the one hand it would be appropriate by the lights of MMT for George to donate all of his (spare) money to the Against Malaria Foundation. But, it would also be appropriate for George to donate only the fraction of his money corresponding to his credence in Singer's utilitarianism, and for him to spend the rest on his dream car.

---

[23] See Ross (2006) for some more detail on this argument, and see Tarsney (2019) for additional implications of the dominance argument elsewhere. In response, MacAskill, Bykvist and Ord's have suggested falling back on prudential, or self-regarding, reasons (2020, 52-3). But, although this helps us avoid demandingness in one sense (MacAskill, Bykvist and Ord 2020, 52-3), in another sense it exacerbates the problem: we would be "prohibited from acting against our interests to a certain degree and obligated to act in accordance with our interests to a certain degree" [removed for blind review].

To us, this seems like a neat compromise. For those, like Singer, with high credence in the view that morality is highly demanding, MMT will also be demanding; for those with low credence in such views, accounting for moral uncertainty via MMT will not make moral life very demanding. In this way, MMT recovers an appropriate degree of moral latitude, and thereby isn't guilty of being overly-demanding.

§5

All of the cases that we have considered in the preceding section involve distributing some continuously divisible resource. However, cases like *Old-Timey Villain* do not have this structure. Instead, in these cases the uncertain agent faces a choice between several different discrete options.

We should understand cases like *Mining Safari* as also having this kind of structure. As table 1 describes, there are only four possible choice-worthiness outcomes for Singer and Kagan's views in *Mining Safari*. So, individuating options by their choice-worthiness differences, there are only four possible options available in *Mining Safari*: block Shaft A; block Shaft B; (partially) block both shafts; block neither.

How should MMT handle cases of this sort? We will consider two possibilities. (While these strike us as the two most promising ways of extending MMT, they need not exhaust the possibilities.)

5.a. Lottery tickets
The first way to extend MMT so as to cover these discrete-choice decision problems is to stipulate that in each discrete choice situation, each theory representative will be endowed with a 'lottery ticket' that gives her a chance—equal to the decision maker's credence in the corresponding moral theory—of determining what the decision maker does in that choice situation.

Before the winner of the lottery is determined, theory representatives can make contracts with each other governing what they will do if they win the lottery. We stipulate that each representative wishes to maximize her expected utility under uncertainty about which representative will win the decision lottery. For instance, consider the decision lottery in *Mining Safari*. In the absence of agreeing to a contract, the Kagan and Singer representatives would block Shaft A and Shaft B, respectively, if they won the decision lottery. The Singer representative's expected utility from a coin toss over these two outcomes is $(0.5 \times 10) + (0.5 \times 20) = 15$, and the Kagan representative's expected utility is $(0.5 \times 10) + (0.5 \times 5) = 7.5$. However, if the representatives agreed not to block either shaft regardless of who wins the lottery, then the Singer and Kagan representatives' utilities would be 16 and 8 respectively (see table 1). Each representative regards this contract as an improvement over the disagreement point. In fact, this contract is the Nash bargaining solution in this scenario. MMT recommends that you should not block either shaft in *Mining Safari*.[24]

---

[24] This approach is quite similar to the one adumbrated by Newberry and Ord (2021, 8-9). Newberry and Ord suggest a bargaining approach inspired by analogy with a parliament that uses non-deterministic "proportional chances voting." However, Newberry and Ord do not discuss which formal model of bargaining they think might be appropriate here, as we do in §4.b above.

What about *Old-Timey Villain*? In this case, the decision maker's two representatives do not have anything to gain by making contracts with each other. The first representative just wants to maximize the probability that the decision maker shoves George into the trolley's path, and the second representative just wants to minimize this probability. Thus, in MMT's economic model of *Old-Timey Villain*, the representatives will not agree to any contracts. Under this decision lottery without contracts there is a 10% chance that the decision maker will shove George into the trolley's path, and a 90% chance that she will do nothing.

Randomizing 10-90 over whether or not you should shove George into the path of the trolley will leave a bad taste in some people's mouths, such as (Greaves and Cotton-Barratt 2023, §7.2; Newberry and Ord 2021, 8). They will reply: which actions are super-subjectively permissible when deciding under conditions of moral uncertainty should presumably not depend on the outcomes of random processes in this manner. Moreover, they will argue that it is implausible to suppose that if the lottery resolves in favour of the theory in which the decision maker has 10% credence, then it would be appropriate to kill George (despite the fact that the decision maker has 90% credence in this action being morally reprehensible). In our experience, different people can have sharply divergent intuitions on the plausibility of randomizing in cases like *Old-Timey Villain*. (For instance, one of the present authors regards it as intuitively plausible; but another of us regards it as wildly unattractive.)

5.b. Eliminating randomization
A second way to extend MMT to cover discrete-choice decision problems is to concede that in cases like *Old-Timey Villain*, MMT's appropriateness prescriptions should come apart somewhat from the output of MMT's economic model. In cases like *Mining Safari* where the decision maker's representatives in MMT's economic model agree to a contract which selects a determinate course of action regardless of which representative wins the decision lottery, this extension of MMT recommends that the decision maker should follow this course of action. However, in cases like *Old-Timey Villain* where the representatives in MMT's economic model cannot agree to a contract which selects a determinate course of action regardless of who wins the decision lottery, this second extension of MMT instead recommends that the decision maker should simply perform the course of action that has the greatest probability of being selected by the representatives in MMT's economic model after the decision lottery occurs. For instance, in *Old-Timey Villain*, the decision maker should *not* push George onto the tracks. In cases where two or more options are tied for the greatest probability, all of the tied options are super-subjectively permissible according to this second extension of MMT.

Under this extension of MMT, the representatives of each moral theory in which one has credence will continue to bargain within our economic model *as if* whoever wins the lottery will in fact get to decide (subject, of course, to any contracts which she has agreed to) which option is chosen by the decision maker. However, this extension of MMT's prescriptions *diverge* from that outcome of the economic model in cases where the lottery still leaves something to chance. The decision maker (super-subjectively) should select the option that has the greatest probability

of being selected by the lottery given the contracts that have been negotiated by the theory-representatives.[25]

One might worry that this decision to partially divorce MMT's recommendations from the outputs of its economic model is *ad hoc*, or theoretically undermotivated. On the contrary, however, we think that this decision can be theoretically motivated by thinking about the fundamental purpose of a theory of appropriate action under conditions of moral uncertainty. The purpose of a theory like this is to recommend some concrete plan of action to the morally uncertain decision maker. The first extension of MMT does not fulfill this purpose, unlike the second.

We leave it open as to which extension is more plausible. Each has its merits, and opinions will doubtless be mixed. For our purposes here, it is enough to have shown that MMT *can be* extended to handle cases where the uncertain agent faces a choice between several different discrete options.

§6

In this section, we address the problem of small worlds.

Using a Nash bargaining model to handle the problem of moral uncertainty has previously been discussed by Greaves and Cotton-Barratt (2023). They "tentatively conclude that while the bargaining approach is interesting, in the end it seems inferior to at least one version of the standard 'maximize expected choice-worthiness' approach" (Greaves and Cotton-Barratt 2023, 27).

There are several important differences between MMT and Greaves and Cotton-Barratt's proposals. For instance, MMT's disagreement point in resource division cases is different from any of the potential disagreement points proposed by Greaves and Cotton-Barratt (2023, §3.1). This difference can be traced to the fact that MMT is specifically inspired by the kind of bargaining that takes place in free markets with well-defined property rights over goods and labor (see §4 above). At the disagreement point, each representative is endowed with her fair share of property rights over the decision maker's resources.[26]

Hence, some of Greaves and Cotton-Barratt's objections to the bargaining proposals that they consider in their paper are inapplicable to MMT. For instance, Greaves and Cotton-Barratt worry that the proposals they consider will sometimes recommend randomizing over several possible options instead of choosing a particular course of action with certainty (2023, §7.2). However, as previously noted, the extension of MMT that we considered in §5.b above avoids this problem.

Nonetheless, one of Greaves and Cotton-Barratt's main objections to Nash bargaining approaches is relevant to MMT. They present readers with the following scenario:

> *Two Binary Choices.* Jenny faces two independent binary choices. She
> can either kill one person, or let two die; and she can either donate a

---

[25] This approach is similar to the one adumbrated by Newberry and Ord (2021, 8-9). For a critical discussion of Newberry and Ord, see [redacted for blind review].

[26] Also, MMT uses the symmetric version of the Nash bargaining solution, whereas Greaves and Cotton-Barratt's proposals use the asymmetric version. This difference is unimportant for our purposes here.

fixed philanthropic budget of (say) \$1m to support homeless people, or to mitigate extinction risk. Her credence is split equally between two moral theories. Jenny has 50% credence in a total utilitarian theory T_1, according to which it is (relatively speaking) slightly better to kill one than to let two die, but much better to direct the resources to extinction risk mitigation than to homeless support. And she has 50% credence in a common-sense moral theory T_2, according to which it is (relatively speaking) slightly better to direct resources to homeless support than to extinction risk mitigation, but much worse to kill one than to let two die (2023, 15).

One way of modelling *Two Binary Choices* is to treat it as a single choice situation, in which the agent has four options:

(1) Kill one and support the homeless
(2) Kill one and fund extinction risk mitigation
(3) Let two die and support the homeless
(4) Let two die and fund extinction risk mitigation

Call this the *grander-world model* of *Two Binary Choices*. As Greaves and Cotton-Barratt show, the Nash bargaining solution in the grander-word model is (4): let two die and fund extinction risk mitigation (2023, §6).

An alternative way of modelling *Two Binary Choices* is to treat it as two staggered choice situations, one happening right after the other. In the first choice situation, Jenny has to decide whether to kill the one or to let the two die. In the second choice situation, Jane has to decide whether to support the homeless or to fund extinction risk mitigation. Call this the *smaller-worlds model* of *Two Binary Choices*. The Nash bargaining solution for each of these two choice situations in the smaller-worlds model is 50-50 randomization over the two available options. Thus, on the smaller-worlds model, MMT implies that killing the one and letting the two die could *both* in principle be permissible for Jenny (cf. §5 above), and likewise that supporting the homeless and funding existential risk mitigation could *both* in principle be permissible (likewise, cf. §5 above). Thus, MMT implies that *all* of the four possible combinations (1) – (4) could in principle be permissible for Jenny in *Two Binary Choices*.

Greaves and Cotton-Barratt call this the *problem of small worlds* (2023, §6). "It can make a significant difference to the verdict of the Nash approach whether one chooses a smaller- or a grander-world model of one's decision problem" (Greaves and Cotton-Barratt 2023, 27). According to Greaves and Cotton-Barratt, "this is problematic, since any such choice (short of the impractical maximally grand-world model) seems arbitrary" (2023, 27). The options assumed to be open to an agent in a "maximally grand-world model" are presumably complete plans of action for the remainder of the agent's lifetime, specifying how the agent would act under every possible empirical contingency.

However, we think that Greaves and Cotton-Barratt are too quick to conclude that a rule for deciding on some privileged small-world model of an agent's circumstances must be

"arbitrary". Reflecting on the circumstances in which theories of appropriate action under moral uncertainty like MMT are designed to be applied suggests a principled and intuitively attractive principle for deciding how 'grand' our model of an agent's option-set should be made in any particular choice situation.

Theories of appropriate action like MMT are designed to be applied in circumstances where an agent is trying to decide what to do in light of her moral uncertainty. Hence, it makes sense to think of an agent's set of options in any particular choice situation as the set of plans of action that she is capable of *deciding* to perform in that choice situation.[27] Call this approach *decisionism*. It is beyond the scope of this paper to propose necessary and sufficient conditions for 'being able to decide to ϕ' (cf. Hedden 2012, §7; Shepherd 2015). However, on any plausible rendering of these conditions, decisionism will almost always select worlds smaller than the maximally grand world, given that no human agent has the cognitive capacities one would require in order to decide now on one particular complete plan of action for the remainder of one's lifetime (apart from agents who know they are at death's door). On the other hand, almost all human agents have the cognitive capacity to decide to perform any of the four possible options (1) – (4) from Greaves and Cotton-Barratt's grander-world model of *Two Binary Choices*. Thus, MMT plus decisionism implies that (4) is the only permissible option in *Two Binary Choices*. This strikes us as the right result.

Although the precise details of decisionism will depend on exactly how one analyses 'being able to decide to ϕ,' we are cautiously optimistic that decisionism (or something like it) can provide a principled and intuitively attractive response to Greaves and Cotton-Barratt's problem of small worlds. Despite their objections, the Nash bargaining approach to moral uncertainty remains a viable option.[28]

§7

In this paper, we have motivated a bargaining approach to decision making under moral uncertainty. The specific theory we presented in §4, MMT, captures our intuitions about the appropriateness of splitting one's donations in proportion to one's credences in various moral theories, and provides a satisfying explanation for when and why departures from the proportional response are appropriate.

As was discussed in §4, capturing Proportionality is not the only advantage MMT has over traditional views on appropriate choice under conditions of moral uncertainty. MMT successfully circumvents a number of potential pitfalls that have divided the field, such as intertheoretic comparability. But for all that, we have only scratched the surface of the bargaining-based approach to moral uncertainty. MMT is a jumping off point, meant to illustrate both the viability and appeal of the bargaining approach more generally, and hopefully to spark more interest in this project.

---

[27] This approach is inspired by (Hedden 2012).
[28] It is also worth noting here that the Maximize Expected *Normalized* Choice-Worthiness extension of MEC designed to cover cases of intertheoretic unit-incomparability also suffers from a problem of small worlds [removed for blind review].

Bibliography

Beckstead, Nick and Thomas, Teruji (2023). A Paradox for Tiny Probabilities and Enormous Values, *Noûs* 00: 1-25. https://onlinelibrary.wiley.com/doi/10.1111/nous.12462

Benatar, David (2006). *Better Never to Have Been: The Harm of Coming into Existence*. Oxford: Oxford University Press.

Bennett, Jonathan (1978). On Maximizing Happiness. In R. I. Sikora and B. Barry (eds.), *Obligations to Future Generations* (pp. 61-73). Philadelphia: Temple University Press.

Binmore, Ken, Rubinstein, Ariel and Wolinsky, Asher (1986). The Nash Bargaining Solution in Economic Modelling, *RAND Journal of Economics* 17(2): 176-188.

Bostrom, Nick (2011). Infinite Ethics, *Analysis and Metaphysics* 10: 9-59.

Briggs, Ray A. (2019). Normative Theories of Rational Choice: Expected Utility. In Edward N. Zalta (ed.), *The Stanford Encyclopaedia of Philosophy*, Fall 2019. https://plato.stanford.edu/archives/fall2019/entries/rationality-normative-utility/

Broome, John (2012). *Climate Matters: Ethics in a Warming World*. New York: W. W. Norton.

Clark, Matthew and Pummer, Theron (2019). Each-We Dilemmas and Effective Altruism, *Journal of Practical Ethics* 7(1): 24-32.

Field, Claire (2019). Recklessness and Uncertainty: Jackson Cases and Merely Apparent Asymmetry, *Journal of Moral Philosophy* 16(4): 391-413.

Gracely, Edward J. (1996). On the Noncomparability of Judgements Made by Different Ethical Theories, *Metaphilosophy* 27(3): 327-332.

Greaves, Hilary and Cotton-Barratt, Owen (2023). A Bargaining-Theoretic Approach to Moral Uncertainty. *Global Priorities Institute Working Paper No. 2-2023, forthcoming in the Journal of Moral Philosophy.* https://globalprioritiesinstitute.org/a-bargaining-theoretic-approach-to-moral-uncertainty/

Greaves, Hilary, MacAskill, William, Mogensen, Andreas and Thomas, Teruji (2022). On the Desire to Make a Difference. *Global Priorities Institute Working Paper No. 16-2022*. https://globalprioritiesinstitute.org/on-the-desire-to-make-a-difference-hilary-greaves-william-macaskill-andreas-mogensen-and-teruji-thomas-global-priorities-institute-university-of-oxford/

Greaves, Hilary and Ord, Toby (2017). Moral Uncertainty About Population Axiology, *Journal of Ethics and Social Philosophy* 12(2): 135-167.

Gustafsson, Johan (2022). Second Thoughts About My Favourite Theory, *Pacific Philosophical Quarterly* 103(3): 448-470.

Gustafsson, Johan and Torpman, Olle (2014). In Defence of My Favourite Theory, *Pacific Philosophical Quarterly* 95(2): 159-174.

Harman, Elizabeth (2009). Critical Study: David Benatar's Better Never to Have Been: The Harm of Coming into Existence, *Noûs* 43(4): 776-785.

——————— (2014). The Irrelevance of Moral Uncertainty. In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics, Volume 10* (pp. 53-79). Oxford: Oxford University Press.

Hedden, Brian (2012). Options and the Subjective Ought, *Philosophical Studies* 158(2): 343-360.

——————— (2016). Does MITE Make Right? In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics, Volume 11* (pp. 102-128). Oxford: Oxford University Press.

Kagan, Shelly (2019). *How to Count Animals, More or Less.* Oxford: Oxford University Press.

MacAskill, William (2016). Normative Uncertainty as a Voting Problem, *Mind* 125(500): 967-1004.

MacAskill, William, Bykvist, Krister and Ord, Toby (2020). *Moral Uncertainty*. Oxford: Oxford University Press.

MacAskill, William and Ord, Toby (2020). Why Maximize Expected Choice-Worthiness?, *Noûs* 54(2): 327-353.

Mogensen, Andreas L. (2016). Should We Prevent Optimific Wrongs?, *Utilitas* 28(2): 215-226.

Muñoz, Daniel and Spencer, Jack (2021). Knowledge of Objective 'Oughts': Monotonicity and the New Miners Puzzle, *Philosophy and Phenomenological Research* 103(1): 77-91.

Nash, John F. Jr. (1950). The Bargaining Problem, *Econometrica* 18(2): 155-162.

Newberry, Toby and Ord, Toby (2021). The Parliamentary Approach to Moral Uncertainty. *FHI Technical Report No. 2021-2*. https://www.fhi.ox.ac.uk/wp-content/uploads/2021/06/Parliamentary-Approach-to-Moral-Uncertainty.pdf

Ord, Toby (2015). Moral Trade, *Ethics* 126(1): 118-138.

Pallies, Daniel (forthcoming). Pessimism and Procreation, *Philosophy and Phenomenological Research* 00: 1-32*. https://onlinelibrary.wiley.com/doi/abs/10.1111/phpr.13007

Parfit, Derek (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Pummer, Theron (2016). Whether and Where to Give, *Philosophy & Public Affairs* 44(1): 77-95.

——————— (2021). Impermissible yet Praiseworthy, *Ethics* 131(4): 697-726.

——————— (2023). *The Rules of Rescue: Cost, Distance, and Effective Altruism*. Oxford: Oxford University Press.

Roberts, Melinda (2003). Is the Person-Affecting Intuition Paradoxical?, *Theory and Decision* 55(1): 1-44.

Ross, Jacob (2006). Rejecting Ethical Deflationism, *Ethics* 116(4): 742-768.

Sepielli, Andrew (2010). *'Along an Imperfectly Lighted Path': Practical Rationality and Normative Uncertainty*. PhD dissertation. New Brunswick, NJ: Department of Philosophy, Rutgers University.

Shepherd, Joshua (2015). Deciding as Intentional Action: Control Over Decisions, *Australasian Journal of Philosophy* 93(2): 335-351.

Sinclair, Thomas (2018). Are We Conditionally Obligated to be Effective Altruists?, *Philosophy & Public Affairs* 46(1): 36-59.

Singer, Peter (1972). Famine, Affluence, and Morality, *Philosophy & Public Affairs* 1(3): 229-243.

——————— (2009). *Animal Liberation*. New York: HarperCollins.

Sung, Leora (2022). Never Just Save the Few, *Utilitas* 34(3): 275-288.

Tarsney, Christian J. (2018). Moral Uncertainty for Deontologists, *Ethical Theory and Moral Practice* 21(3): 505-520.

——————— (2019). Rejecting Supererogationism, *Pacific Philosophical Quarterly* 100(2): 599-623.

——————— (2021). Vive la Différence? Structural Diversity as a Challenge for Metanormative Theories, *Ethics* 131(2): 151-182.

Temkin, Larry (2022). *Being Good in a World of Need*. Oxford: Oxford University Press.

Weatherson, Brian (2019). *Normative Externalism*. Oxford: Oxford University Press.

Wilkinson, Hayden (2022). In Defense of Fanaticism, *Ethics* 132(2): 445-477.