Moral Uncertainty, Pure Justifiers, and Agent-Centred Options

Patrick Kaczmarek and Harry R. Lloyd*

Penultimate draft of paper forthcoming in Australasian Journal of Philosophy.

1 Introduction

We often have to make moral decisions under empirical uncertainty. For instance, consider the following well-worn thought experiment:

> *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, both shafts will partially flood and 10 miners will die. You also know that if you block the shaft where the miners are, then you will save all 100 of them. However, if you block the empty shaft, then the other shaft will flood completely, killing all 100 miners. Finally, your evidence doesn't tell you whether the miners are in Shaft A or B. For you, it's a 50-50 guess.[1]

What should you do?

There is a clear sense in which you should block Shaft A; that would be objectively best. Moreover, there is also a clear sense in which you ought not to leave both shafts unblocked. After all, you know that if the miners are trapped in Shaft A, then you objectively should block Shaft A. Furthermore, you also know that if the miners

*Special thanks to Stephen Darwall, Paul Forrester, Dan Greco, Daniel Muñoz, two blind reviewers, and an editor of the Australasian Journal of Philosophy for fabulous comments. Thanks also to Andreas Mogensen for pointing us in the right direction.

[1]The puzzle is original to (Parfit, 1988; Regan, 1980). Here, we reproduce the version from (Muñoz & Spencer, 2021, p. 78).

are trapped in Shaft B, then you objectively should block shaft B. Thus, leaving both shafts unblocked is sure to be a wrong thing to do.

However, your doxastic attitudes being what they are, there is also a sense in which it seems reckless to block either Shaft A or B. Because the location of the miners is a mystery to you, you are as likely to kill these 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 & Spencer, 2021, p. 79).

Rather, you should act in the way that would be expectably-best. As Derek Parfit (2011, p. 160) painstakingly explains,

> To decide which of our possible acts would make things go *expectably-best*, we take into account both how good the effects of the different possible acts might be, and the probabilities, given our beliefs or the available evidence, that these acts would have these effects. …. In [*Miners*], for example, if we closed either [Shaft A] or [Shaft B], the expectable number of lives would be 100 multiplied by a chance of one in two, or by 0.5. This number would be 50. If we closed [neither shaft], this expectable number would be 90, since this act would be certain to save 90 lives.

Few disagree.

More controversially, some philosophers—notably William MacAskill, Krister Bykvist, and Toby Ord—claim that a similar decision procedure should be used to determine what it is appropriate to do when one is uncertain not (merely) about which empirical state of affairs obtains, but (also) about which normative theory is

correct.[2,3] On this view, what it is appropriate to do when one is morally uncertain depends upon two things: (i) the credence that the decision-maker assigns to various different moral theories; and (ii) the degree of choiceworthiness that those moral theories assign to different options (MacAskill & Ord, 2020, p. 336).[4] The 'choiceworthiness' of any given option represents the all-things-considered strength of one's reasons for choosing that option (MacAskill & Ord, 2020, p. 329).

According to this **Maximise Expected Choiceworthiness ('MEC')** theory,

> 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 (MacAskill & Ord, 2020, p. 338).

There are several potential objections to MEC. In this paper, we will discuss one particular class of objections, *viz.* those occasioned by first-order moral theories according to which suboptimal options can sometimes nonetheless be permissible.

Consider:

> *Self Sacrifice.* You spot a runaway trolley barrelling towards five innocent strangers who cannot get out of its way in time. These strangers will die unless you place your legs in the trolley's path, which would

[2]One important argument in favour of approaching moral uncertainty this way is the *argument from analogy*. It runs as follows:

1. Moral uncertainty should be handled analogously to empirical uncertainty.
2. The right way to handle empirical uncertainty is to *maximise expected value*.

Therefore

3. The right way to handle moral uncertainty is to maximise expected choiceworthiness.

MacAskill et al. (2020, pp. 47–8) employ the argument from analogy in their defense of Maximise Expected Choiceworthiness, arguing that since "Expected utility theory is the standard account of how to handle empirical uncertainty probabilities ... maximizing expected choiceworthiness should be the standard account of how to handle moral uncertainty". Similarly, Christian Tarsney (2021, p. 172) maintains that treating moral and empirical uncertainty "differently when we are not forced to is at least prima facie inelegant and undermotivated" (cf. Sepielli, 2010, pp. 75–8). For criticism of the argument from analogy, see Kaczmarek et al. (2023, §3).

[3]Other philosophers deny there are subjective norms that guide our actions under conditions of moral uncertainty; see esp. Harman (2014) and Weatherson (2019). We won't engage with those arguments here: our question is only whether these subjective norms, if there are any, make space for agent-centred options.

[4]What it is appropriate to do may also be sensitive to any dependencies between your moral and descriptive uncertainties (Podgorski, 2020).

destroy your legs but also bring the trolley to a grinding halt before it reaches the five.

Suppose that you are torn between two moral theories. On the one hand, you are confident in the truth of a certain precisification of commonsense morality. According to this commonsense moral theory, you are not required to sacrifice your legs. However, it is nonetheless permissible for you to venture beyond the call of duty in order to save the five. In other words, it is supererogatory for you to sacrifice your legs in *Self Sacrifice*. On the other hand, you have at least some credence in the demanding view that it would be seriously wrong for to refrain from saving the five—in other words, the needs of the many outweigh your own.

How does MEC imply that one should behave in *Self Sacrifice*? Well, according to the demanding view, sacrificing yourself is clearly much more choiceworthy than failing to do so. But how do these actions compare in choiceworthiness according to commonsense morality? There are three obvious possibilities to consider.

The first possibility is that sacrificing your legs is more choiceworthy than failing to do so according to commonsense morality as well as according to the demanding view. According to some possible views about the nature of supererogation, some supererogatory options are more strongly supported by one's reasons than their merely permissible alternative are; for instance, see Horgan and Timmons (2017) for a view of this sort. If sacrificing your legs in *Self Sacrifice* is an option of this kind, then MEC clearly implies that it is the uniquely appropriate option in this choice situation.

The second possibility is that sacrificing your legs is *less* choiceworthy than failing to do so according to commonsense morality. Recall that the choiceworthiness of any given option represents the *all-things-considered* strength of one's reasons for choosing that option. Now, as part of common sense, you presumably have strong non-moral prudential reasons to protect your own body from harm (Scheffler, 1982; Slote, 1984). On balance, these self-regarding reasons against sacrificing your legs might be stronger than the moral reasons in favour of it. Hence, sacrificing your legs might be less choiceworthy than failing to do so. In this case, it is not clear which action MEC selects as the most appropriate. If sacrificing your legs is sufficiently unchoiceworthy according to common sense, then doing nothing will maximise expected choiceworthiness. On the other hand, if sacrificing your legs is only slightly unchoiceworthy according to common sense, then sacrificing your legs might still maximise expected choiceworthiness. Finally, if—as a matter of coincidence—sacrificing your legs and doing nothing both happen to have exactly the same expected choiceworthiness, then MEC will imply that both of these

actions are appropriate in *Self Sacrifice*.

The third possibility is that sacrificing your legs and failing to do so are equal in choiceworthiness according to commonsense morality. Under this assumption, sacrificing your legs clearly maximises expected choiceworthiness (cf. Ross, 2006). Thus, MEC implies that it is the uniquely appropriate option in *Self Sacrifice*.

One important thing to notice about all three of these possible cases is that MEC does not allow for the category of supererogatory options as part of its verdicts about appropriateness.[5] Sometimes, as a matter of coincidence, two or more options might be tied for maximal expected choiceworthiness. In all other cases, however, MEC always implies that only one possible option is appropriate. For instance, in *Self Sacrifice* MEC implies either that you subjectively ought to sacrifice your legs or that you subjectively ought not to. We find this result counterintuitive. It is more plausible to suppose that sacrificing your legs and doing nothing are both appropriate options in *Self Sacrifice*.

We have supposed that you are confident in the truth of commonsense morality. Furthermore, we can assume that commonsense morality is a theory that regards the existence of agent-centred options as a central and important feature of morality. Deontological moral philosophers have argued that this moral latitude reflects a fundamental feature of our moral status, *viz.* that persons matter unconditionally (Lazar, 2019, pp. 89ff). As F. M. Kamm (1992, pp. 358–9) forcefully puts it, persons are "ends-in-themselves, having a point even if they do not serve best consequences" (cf. Chappell, 2015). Depriving a person of (nearly all) moral latitude whenever she is morally uncertain arguably diminishes her status as an important moral creature (Barry & Tomlin, 2016).

Importantly, the key idea here is not that persons are sometimes rationally *required* to prioritise their own interests over those of others. Rather, the key idea is that persons at least sometimes have the *choice* to prioritise their own interests over those of others. Advocates of MEC who claim that it is appropriate to do nothing in *Self Sacrifice* because your commonsense prudential concerns outweigh your commonsense moral concerns in this scenario (Lockhart, 2000; MacAskill, 2019; MacAskill et al., 2020) seem to misconstrue the all-things-considered structure of

[5] One further possibility which we have not mentioned yet is that the choiceworthiness of sacrificing your legs and failing to do so might be 'on a par' (Parfit, 2011, pp. 137–41); see also Muñoz (2021) and references therein. 'Parity' is a term used by philosophers like Ruth Chang (2002) to refer to a certain relationship of incommensurability. In this case, however, it will be impossible to represent the choiceworthiness of each of these two options in terms of a single number, so standardly-formulated MEC seems to be simply inapplicable (cf. §§4-5 below). We will relax the assumption that choiceworthiness is unidimensional in §2ff below.

your reasons according to common sense (Sung, ms); see also (Hedden, 2016, §5.2.2). Neither sacrificing your legs nor doing nothing is uniquely favoured by our commonsense reasons in *Self Sacrifice*. (We elucidate this idea in greater detail in §2 below.)

In certain cases, such as *Self Sacrifice*, we think it is implausible for MEC to imply that only one option is appropriate if the decision maker has high credence in theories that afford the decision-maker an agent-centred prerogative. Furthermore, MEC's general approach to agent-centred prerogatives also strikes us as implausible, even when we abstract away from the details of any particular cases. Latitude is only ever a matter of coincidence on MEC; in all possible choice situations other than those in which two or more options happen to be tied for maximal expected choiceworthiness, MEC implies that only one possible option is uniquely appropriate. MEC has this implication regardless of the decision maker's level of credence in theories that endorse agent-centred prerogatives.

This strikes us as an unattractive feature for a theory of appropriateness. A better theory of appropriateness would be more sensitive to the decision maker's credence in theories that endorse agent-centred prerogatives. To the extent that an agent has high credence in theories that regard the existence of agent-centred prerogatives as an important feature of the normative domain, a good theory of decision making under moral uncertainty should allow for the category of supererogatory options as part of its verdicts about appropriateness. By contrast, to the extent that an agent has low or zero credence in agent-centred prerogatives, a good theory of decision making under moral uncertainty can legitimately select a single option as uniquely appropriate in all or most choice situations.

In this paper, we will develop and defend a modified version of MEC that allows for agent-centred prerogatives and supererogation as part of its verdicts about appropriateness.[6] We begin, in §2, by introducing the distinction between *requiring* and *justifying reasons*. We also introduce Daniel Muñoz's rights-based account of prerogatives as a convenient case study. Then, in §3, we will introduce and explain our new **Expected Balance of Reasons ('EBR')** theory of appropriate choice under conditions of moral uncertainty. Across §§4-5, we will argue that EBR overall compares favourably with its rivals. §6 wraps everything up, and draws out the practical implications of EBR for charitable activities.

[6]One extant alternative to MEC that already allows for this is **My Favourite Theory** (Gracely, 1996; Gustafsson & Torpman, 2014). Unfortunately, however, this approach to handling decisions when morally uncertain faces several decisive objections (Greaves & Ord, 2017; Gustafsson, 2022; MacAskill et al., 2020).

2 Two kinds of reasons

MEC implicitly assumes that choiceworthiness—the all-things-considered strength of one's reasons for choosing any given option—is a *unidimensional* variable. In other words, MEC assumes that the choiceworthiness value of any given option can always be measured by a single number (Tarsney, 2021). By contrast, we want to suggest that choiceworthiness is a *multidimensional* variable. In particular, we want to draw a distinction between (i) the all-things-considered requiring strength of one's reasons for choosing any given option, and (ii) their all-things-considered justifying strength.[7]

In brief: the justifying strength of some reason in favour of an option measures the extent to which that reason *pro tanto* counts in favour of the option being *permissible*. By contrast, the requiring strength of some reason in favour of an option measures the extent to which that reason *pro tanto* counts in favour of the option being *required*.[8] An option A is in fact permissible iff

> the all-things-considered justifying strength of one's reasons in favour of A is greater than or equal to the all-things-considered requiring strength of one's reasons in favour of any alternative option B.

On the flipside, an option A is in fact required iff

> the all-things-considered requiring strength of one's reasons in favour of A is greater than the all-things-considered justifying strength of one's reasons in favour of any alternative option B.

It will be helpful to illustrate this distinction using Daniel Muñoz's (2021) rights-based account of prerogatives as a case study. Although we use Muñoz's view as our case study in this paper, his is just one of many theories according to which the justifying strengths of some reasons can come apart from their requiring strengths; including Campbell and Kaczmarek (ms), Hurka and Shubert (2012), Mogensen (2019), Muñoz and Pummer (2022), Pummer (2023), and Thomas (2022). We focus on

[7] Rebelling against the old fashion that reasons exclusively issue *pro tanto* requirements, philosophers are increasingly adopting the position that reasons can vary on at least two dimensions with respect to their normative strength (Gert, 2004; Kamm, 1985; Lazar, 2013). See Little and Macnamara (2021) for a survey of the landscape.

[8] Granted that whatever is required must also be permitted, the justifying strength of the reasons in favour of some option must always be at least as great as their requiring strength.

Muñoz's view in particular only for sake of concreteness. We also want to emphasise that we do not argue in favour of Muñoz's view in this paper. We simply claim that it is a coherent first-order moral theory in which a reasonable decision maker might have some positive credence.

We can unpack Muñoz's proposal in the context of *Self Sacrifice*. Deontologists (tend to) have two things to say about cases of this sort. First, it would be seriously wrong for me to shove you into the path of the runaway trolley. Second, you are not obligated to sacrifice your legs, although you are permitted to do so should you wish. In an attempt to unify these two strands of the deontological response (respectively: side-constraints and agent-centred prerogatives), Muñoz proposes that these two exceptions to promoting the greater good are both grounded in *claims rights* held by somebody against somebody.

First, your rights against harm restrict my choices. All else being equal, it would be gravely wrong for me to use your body to stop the trolley without your consent, even if this really would be for the best. So far, so familiar.

Less familiarly, however, Muñoz (2021, p. 609) also proposes that we have the "same basic rights against ourselves as we do against others, rights that proscribe harm and bodily intrusion, and that are waived via consent". Just like I need your consent to permissibly use your body, you too require your own consent to do this. However, since you obviously consent to actions that you deliberately take, you can never intentionally violate your own rights against harm. If you were to place your legs on the tracks, then you would waive the right in the nick of time. And if you don't place your legs on the tracks, then you can lean on this self-directed right in defending your refusal. In this way, Muñoz derives agent-centred prerogatives from waivable rights against oneself.

The idea that claim rights are self-other symmetric is not original to Muñoz; it has been discussed elsewhere, such as in work of Paul Hurley (1995) and Shelly Kagan (1989). What *is* new about Muñoz's proposal is the further assertion that a waivable right against oneself does not generate requiring reasons (cf. Schofield, 2021).

On the orthodox, unidimensional view of practical rationality, reasons are considerations upon which it would be "wrong not to act on in the absence of any opposition" (Dancy, 2004, p. 92). Left unobstructed, a *pro tanto* reason gives way to an all-things-considered deontic ought. According to Muñoz, however, waivable rights against myself do not behave in this fashion. For instance, my claim right against myself over my body does not *pro tanto* count in favour of requiring me to protect my legs (Muñoz, 2021, pp. 615–6). In fact, it is permissible for me to sacrifice my legs even if doing so will only prevent somebody else from losing a finger. This

is foolish, perhaps; but not prohibited.

Thus, according to Muñoz waivable claim rights against myself are pure justifiers. They count in my defense even though they do not constrain my practical deliberation. In his own words,

> There is more to morality than acting on sound deliberation; we must also *defend* our actions when the moral community comes along demanding better. Imagine that I go up to Amanda and say, "How dare you not crush your arm! Bert is in dire need, and you have most reason to save him!" She might defend herself by leaning on her rights. "But it's *my* body," she could insist, "and I'm not willing to harm it." Here Amanda is not making excuses (as if to say, "Give me a break—I'm biased"). Nor is the point that she has special reasons to favor herself. The point is that she doesn't *owe* me a reason. Her rights allow her to act against the balance of reasons; they make it defensible to do less than best. [...] Her right evaporates when acted against, but not when leaned on as a justification. "I have a right that I'm unwilling to waive" is a poor [requiring] reason but, in this case, a decent defense (Muñoz, 2021, p. 617).

We can understand Muñoz-style prerogatives as strong yet purely justifying reasons. Your claim right against yourself does not *pro tanto* count in favour of requiring you to keep your legs. Of course, you plausibly have a prudential requiring reason to prevent your legs from being painfully flattened. But the requiring strength of this prudential reason is plausibly weaker than the requiring strength of the moral reason that you have to rescue the greater number. Thus, the balance of your requiring reasons favours you sacrificing your legs in order to save the five. Fortunately, however, the pure justifying force of the reason in favour of keeping your legs supplied by your claim right against yourself is strong enough to permit you to protect your legs—at least when there are only five people trapped on the tracks and in danger of being run over. This justifying reason in some sense 'defuses' the requiring force of your reasons in favour of saving the five (Gert, 2004).

Exactly how strong is the justifying reason in favour of protecting your legs generated by the claim right that you hold against yourself? How does the justifying strength of this reason compare to the requiring strength of your reasons in favour of saving the five? To answer these questions, we should consider how many people would need to stand in the path of the trolley in order for you to be obligated to sacrifice your legs. If this number is greater than 100 (say), then the justifying reason

in favour of protecting your legs is clearly far stronger than the requiring reason in favour of saving the five. By contrast, if this number is six (say), then the justifying reason in favour of protecting your legs is clearly only slightly stronger than the requiring reason in favour of saving the five. In what follows, we will assume—quite conservatively—that only ten people would need to stand in the path of the trolley in order for you to be obligated to sacrifice your legs.

We can now represent the all-things-considered strengths of your reasons for and against saving the five in this choice situation in a 2×2 matrix of real numbers. In particular, we can suppose that:

Muñoz	**Requiring strength**	**Justifying strength**
Refuse	$1/2$	10
Rescue	5	5

The first number in each of these rows measures the all-things-considered requiring strength of one's reason in favour of the option; the second number measures the all-things-considered justifying strength. On Muñoz's view, your requiring reasons in favour of refusing to rescue are substantially weaker than your requiring reasons in favour of rescuing the five in *Self Sacrifice*. However, you also have a strong justifying reason in favour of refusing to rescue. Because the strength (10) of this justifying reason is greater than the strength (5) of your requiring reasons in favour of rescue, both options are permissible in this choice situation. Furthermore, rescue is the supererogatory option in this scenario, since your requiring reasons favour rescuing as opposed to refusing.

A final question that we need to consider in this section is how one should represent the justifying and requiring strengths of one's reasons according to maximising moral theories like Peter Singer's utilitarianism. According to moral theories such as Singer's, the justifying and requiring aspects of reasons never come apart. An option is permissible in some choice situation iff the all-things-considered reasons in favour of the option have maximal requiring strength in that choice situation. Thus, according to Singer, the justifying strength of one's reasons in favour of any given option must always be equal to their requiring strength. In particular, we can suppose that:

Singer	Requiring strength	Justifying strength
Refuse	1/2	1/2
Rescue	5	5

According to Singer, the justifying strength (1/2) of your reasons to refuse is lesser than the requiring strength (5) of your reasons to rescue. Thus, rescuing is required and refusing is prohibited.

3 EBR

Recall that MEC implicitly assumes that choiceworthiness is unidimensional. By contrast, we have suggested that choiceworthiness has two conceptually-separable dimensions. The first dimension measures the extent to which one's reasons *pro tanto* count in favour of a requirement, whereas the second dimension measures the extent to which one's reasons *pro tanto* count in favour of permissibility.

A natural first-step in modifying MEC in order to accommodate this new multi-dimensional framework is to calculate the expected value of each of our dimensions of choiceworthiness for each possible option. For instance, suppose that your credence is split 50-50 between the Muñoz response and the Singer response to *Self Sacrifice*. Under this assumption, the expected strengths of your justifying and requiring reasons for your two options will be:

Expected	Requiring strength	Justifying strength
Refuse	1/2	5.25
Rescue	5	5

How should we use these results to determine whether each of these options is appropriate in this choice situation? We suggest that one should use exactly the same procedure as one uses to determine at the first-order level which actions are permissible according to any given moral theory.

According to EBR, an option A is *appropriate* iff

> the expected all-things-considered justifying strength of one's reasons in favour of A is at least as great as the expected all-things-considered

> requiring strength of one's reasons in favour of any alternative option B.

Correspondingly, an option A is *uniquely appropriate* iff

> the expected all-things-considered requiring strength of one's reasons in favour of A is greater than the expected all-things-considered justifying strength of one's reasons in favour of any alternative option B.

Thus, according to EBR refusing and rescuing are both appropriate options in *Self Sacrifice*. On the one hand, the expected justifying strength (5.25) of your reasons to refuse is greater than the expected requiring strength (5) of your reasons to rescue. On the other hand, the expected justifying strength (5) of your reasons to rescue is greater than the expected requiring strength (½) of your reasons to refuse. So, both options count as appropriate according to EBR.

However, things would be different if there were more people standing in the path of the trolley. For instance, if there were eight people in danger, then we can suppose that according to Muñoz:

Muñoz	**Requiring strength**	**Justifying strength**
Refuse	½	10
Rescue	8	8

Likewise, according to Singer:

Singer	**Requiring strength**	**Justifying strength**
Refuse	½	½
Rescue	8	8

Under these conditions the expected strengths of your justifying and requiring reasons for these two options will be:

Expected	**Requiring strength**	**Justifying strength**
Refuse	½	5.25
Rescue	8	8

Thus, EBR implies that rescuing is the uniquely appropriate option in this variant choice situation: the expected justifying strength (5.25) of your all-things-considered reasons to refuse is less than the expected requiring strength (8) of your all-things-considered reasons to rescue.

EBR's implications in these kinds of self-sacrificing trolley problems are sensitive to the degree of importance that the 'Muñoz' theory ascribes to your agent-centred prerogatives. We have been considering a version of the theory that regards these prerogatives as fairly unimportant: only ten people would need to stand in the path of the trolley in order for you to be obligated to sacrifice your legs. By contrast, if we assumed that the 'Muñoz' theory regards these prerogatives as much more important, then a hundred people might need to stand in the path of the trolley in order for you to be obligated to sacrifice your legs. Under this assumption, if eight people were in danger, then:

Muñoz	**Requiring strength**	**Justifying strength**
Refuse	1/2	100
Rescue	8	8

Hence, the expected strengths of your justifying and requiring reasons for the two options would be:

Expected	**Requiring strength**	**Justifying strength**
Refuse	1/2	50.25
Rescue	8	8

Thus, EBR implies that refusing and rescuing would both be appropriate options.

In summary: if agent-centred prerogatives are relatively unimportant according to Muñoz's view, then it is uniquely appropriate for you to rescue the eight in this choice situation. But if agent-centred prerogatives are relatively important according to Muñoz's view, then it is also appropriate for you to refuse to rescue the eight in this choice situation. This strikes us as the right result: more robust agent-centred prerogatives should be *eo ipso* more durable under moral uncertainty (*ceteris paribus*).

4 Alternatives to EBR

MEC implicitly assumes that choiceworthiness is unidimensional. This implicit assumption is innocuous in cases where the decision maker has positive credence only in first-order theories like Singer's according to which the justifying and requiring strengths of reasons are always equal to each other. For theories like this, a single 'choiceworthiness' value can represent both the justifying and the requiring strengths of any given reason. (In the remainder of this paper, we ourselves will sometimes use 'choiceworthiness' in this manner to characterise theories like Singer's.)

Moreover, for decision makers who have positive credence only in these kinds of theories, EBR and MEC always deliver exactly the same appropriateness verdicts. According to MEC, some option A is appropriate iff its expected choiceworthiness is greater than or equal to the expected choiceworthiness of any alternative option B. And according to EBR, some option A is uniquely appropriate iff the expected all-things-considered requiring strength of one's reasons in favour of A is greater than the expected all-things-considered justifying strength of one's reasons in favour of any alternative option B. When 'choiceworthiness,' 'justifying strength', and 'requiring strength' have identical values, these two analyses of appropriateness are extensionally equivalent.

What about cases where the decision maker has positive credence in one or more first-order theories like Muñoz's according to which the justifying and requiring strengths of a reason sometimes differ? In these cases, we think that MEC is simply *inapplicable*. According to theories like Muñoz's, the all-things-considered strength of one's reasons for choosing an option cannot always be represented by a single choiceworthiness value. MEC's definition of choiceworthiness is at best ambiguous for these kinds of theories.

Thus, we may regard EBR as a *generalization* of MEC. The cases in which MEC is applicable are a proper subset of the cases in which EBR is applicable; and EBR agrees with MEC in all of the cases where MEC is applicable.

However, EBR is not the only possible generalization of MEC that one might invent to handle cases in which the decision maker has positive credence in one or more theories like Muñoz's. In this section, we consider two natural alternatives to EBR. We will argue that EBR is more attractive than either of these alternative generalizations of MEC.

The first alternative that we will consider is **Maximise Expected Requiring Strength ('MERS')**. According to MERS,

> an option A is appropriate in some choice situation iff A maximises the expected all-things-considered requiring strength of one's reasons in this choice situation.

In *Self Sacrifice*, MERS implies that rescuing the five is uniquely appropriate, given that the expected all-things-considered requiring strength (5) of one's reasons in favour of rescuing the five is greater than the expected all-things-considered requiring strength (½) of one's reasons in favour of refusing to do so.[9]

One problem with MERS is that it makes the relationship between appropriateness and the expected strengths of one's reasons disanalogous to the relationship between permissibility and the actual strengths of one's reasons. Recall that an option is in fact permissible iff the all-things-considered justifying strength of one's reasons in favour of A is greater than or equal to the all-things-considered requiring strength of one's reasons in favour of any alternative option B. By contrast, according to MERS the appropriateness of an option A only depends on the expected requiring strength of one's reasons in favour of A, and does not depend whatsoever upon the expected justifying strength of those reasons. This disanalogy strikes us as *prima facie* unattractive and in need of explanation. It also produces strange implications in certain cases.

For instance, imagine that I face two possible options, A and B. And suppose also that I have positive credence in only two moral theories. According to both of these moral theories, option A and B are both permissible in this choice situation, although both theories also agree that option A is supererogatory, in the sense that my requiring reasons favour option A over option B.

It strikes us as highly plausible to suppose that A and B are both appropriate options to perform in this choice situation. After all, by assumption I am certain that options A and B are both permissible. By contrast, however, MERS implies that option A is uniquely appropriate, since by assumption I am also certain my requiring reasons favour A over B. Thus, option B is inappropriate according to MERS, even though I am certain that it is permissible. This result strikes us as unattractive.

A second alternative to EBR is **Minimise Expected Wrongfulness ('MEW')**. The rough idea behind MEW is that when one is morally uncertain, one should minimise the extent to which one expects to fall short of acting permissibly (cf. Barry & Tomlin, 2016, pp. 906–10). We may define the 'wrongfulness' of any given option

[9] One might regard MERS as a 'tightening' of EBR. Every option that is appropriate according to MERS is also appropriate according to EBR. However, some options that are appropriate according to EBR are not appropriate according to MERS.

in any given choice situation as a measure of the extent to which that option falls short of being permissible in this choice situation. Specifically,

(a) If option A is permissible, then its wrongfulness is zero.

(b) If option A is impermissible, then its wrongfulness is the difference between (i) the all-things-considered justifying strength of one's reasons in favour of A, and (ii) the all-things-considered requiring strength of one's reasons in favour of the alternative option B for which one's requiring reasons are strongest.

For instance, consider *Self Sacrifice.* According to Muñoz's theory, the wrongfulness of rescuing and refusing are both zero, since both options are permissible. By contrast, according to Singer's theory, the wrongfulness of rescuing is zero, but the wrongfulness of refusing is $5 - 1/2 = 4.5$. This reflects the fact that according to Singer the all-things-considered justifying strength of my reasons to refuse is 4.5 units lower than it would need to be in order to make refusing morally permissible in this choice situation.

According to MEW,

> an option A is appropriate in some choice situation iff A minimises expected wrongfulness in this choice situation.

In *Self Sacrifice*, MEW implies that rescuing the five is uniquely appropriate, since the expected wrongfulness (0) of rescuing the five is lower than the expected wrongfulness ($1/2 \times 4.5 = 2.25$) of refusing to do so.

Unlike MERS, MEW implies that if I am certain in the permissibility of a particular option, then that option must also be appropriate. Any option that I am certain is permissible has zero expected wrongfulness, which is the lowest possible value. Unfortunately, however, there are other cases in which MEW has less attractive implications.

For instance, imagine that I face two possible options, A and B. My credence is split 50-50 between the two moral theories T1 and T2. According to T1, my requiring reasons strongly favour option A over option B; but my justifying reasons also strongly support option B. Hence, options A and B are both permissible according to T1. Numerically:

T1	**Requiring strength**	**Justifying strength**
A	20	20
B	0	20

By contrast, according to T2 my requiring and justifying reasons both marginally favour option B over option A. Hence, option B is uniquely permissible according to T2. Numerically:

T2	Requiring strength	Justifying strength
A	10	10
B	11	11

Thus, the wrongfulness of option A is zero according to T1, and 1 according to T2. And the wrongfulness of option B is zero according to both T1 and T2. Hence, MEW implies that option B is uniquely appropriate in this choice situation.

This implication of MEW strikes us as unattractive. The expected requiring strength (15) of my reasons in favour of option A is greater than the expected requiring strength (5.5) of my reasons in favour of option B. Under these conditions, it is implausible to suppose that B but not A is appropriate in this binary choice situation. By contrast, EBR has the attractive implication that A and B are both appropriate options in this choice situation.

We also think that EBR's theoretical structure is more attractive than MEW's. Recall that (by definition) a justifying reason in favour of option A *pro tanto* counts in favour of option A being permissible, and a requiring reason in favour of A *pro tanto* counts in favour of A being required. Closely analogously, EBR claims that having some credence in the existence of some justifying reason in favour of A *pro tanto* counts in favour of A being appropriate, and having some credence in the existence of some requiring reason in favour of A *pro tanto* counts in favour of A being uniquely appropriate. MEW eschews—even more egregiously than MERS—this connection between the first- and second-order *pro tanto* normative significance of requiring and justifying reasons. This disanalogy strikes us as *prima facie* unattractive and in need of explanation.

5 Evaluating EBR

EBR is a generalization of MEC, and thus shares several of MEC's advantages and disadvantages as a theory of appropriate choice under conditions of moral uncertainty. In this section, we discuss several of those advantages and disadvantages of EBR.

One feature that EBR shares with MEC is an openness to 'moral hedging.' For instance, consider a *Miners*-style choice (cf. §1) between three options, A, B, and C. My credence is split 50-50 between the two moral theories T1 and T2. The choiceworthiness values according to T1 and T2 of the options A, B, and C are given by this table:

Choiceworthiness	**T1: 0.5 credence**	**T2: 0.5 credence**
A	100	0
B	90	90
C	0	100

Hence, A's expected choiceworthiness is 50, B's expected choiceworthiness is 90, and C's expected choiceworthiness is 50. B uniquely maximises expected choiceworthiness in this situation. Thus, MEC (and EBR too) implies that B is uniquely appropriate—despite me being certain that B is not the best possible option in this choice situation. This strikes us as a plausible result. In this scenario, I am 100% certain that B is almost as good as the best possible option, but I am highly uncertain about what the best possible option is. Under these circumstances, B is an attractive 'safe' option to select.

In cases where requiring strength comes apart from justifying strength, EBR also allows for some more complicated (and perhaps surprising) forms of moral hedging. For instance, imagine that I face three possible options, D, E, and F. Again, my credence is split 50-50 between theories T1 and T2. According to T1, I have strong requiring reasons in favour of D, moderate justifying reasons in favour of E, and no reasons whatsoever in favour of F. On balance, option D is morally required in this choice situation. Numerically:

T1	**Requiring strength**	**Justifying strength**
D	10	10
E	0	6
F	0	0

By contrast, according to T2 I have no reasons whatsoever in favour of D, moderate justifying reasons in favour of E, and strong requiring reasons in favour of F. On balance, option F is morally required in this choice situation. Numerically:

T2	Requiring strength	Justifying strength
D	0	0
E	0	6
F	10	10

Under these assumptions, the expected strengths of my justifying and requiring reasons for my three options will be:

Expected	Requiring strength	Justifying strength
D	5	5
E	0	6
F	5	5

Thus, EBR implies that D, E, and F are all appropriate options in this choice situation, since the expected justifying strength of my reasons in favour of each option is greater than or equal to the expected requiring strength of my reasons in favour of any other option. In particular, option E is appropriate despite me being certain that E is objectively impermissible in this choice situation.

Although this result is somewhat surprising, we do not think that it is particularly implausible. In this choice situation, I am certain that reasons of moderate strength *pro tanto* count in favour of E of being permissible. By contrast, I only have 50% credence that any kind of reasons count in favour of D—and the same goes for F. Under these circumstances, it strikes us as reasonably plausible to suppose that all three options are appropriate.

Importantly, EBR also supplies additional guidance to the decision maker in this choice situation beyond the verdict that all three options are appropriate. In particular, we may note that D and F are *super-appropriate* in this choice situation—insofar as in expectation I have requiring reasons in favour of choosing D or F that are strong relative to my requiring reasons in favour of choosing E. 'Super-appropriateness' is a second-order analogue of the first-order property of supererogation. Thus, although EBR implies that options D, E, and F are all appropriate, it also implies that there is a sense in which D and F are morally better than E. This strikes us as a plausible response to this choice situation.

Another feature that EBR shares with MEC is *fanaticism*. 'Fanatical' decision theories prefer lotteries with tiny probabilities of arbitrarily high payoffs over guar-

antees of modest payoffs (on fanaticism in general, see Beckstead and Thomas (2023), Cibinel (2023), Russell (2023), and Wilkinson (2022, 2023)). For instance, consider a choice between two options, G and H. I have 99.9% credence in the moral theory T1, and 0.1% credence in the moral theory T2. The choiceworthiness values according to T1 and T2 of the options G and H are given by this table:

Choiceworthiness	**T1: 0.999 credence**	**T2: 0.001 credence**
G	10	10
H	−100	1,000,000

Hence, G's expected choiceworthiness is 10, and H's expected choiceworthiness is 900.1. H uniquely maximises expected choiceworthiness in this choice situation. Thus, MEC (and EBR too) implies that H is uniquely appropriate.

Many of us regard this implication as implausible, and intuit that it would be more appropriate to select option G. After all, in this choice situation I am 99.9% sure that H is highly unchoiceworthy, and 100% certain that G is moderately choiceworthy. Under these circumstances, it seems reckless and uncompromising to prefer H over G.

There are several possible responses to this fanaticism objection to MEC and EBR. One possible response is to argue that rejecting fanaticism has implications that are even more implausible than those of fanaticism itself. For instance, Wilkinson (2023) argues that any decision theory which rejects fanaticism must imply that one's decisions should sometimes be sensitive to one's level of uncertainty about the amount of moral value that was realised in Ancient Egypt. Furthermore, Beckstead and Thomas (2023) argue that non-fanatical decision theories must either (i) "permit passing up an arbitrarily large potential gain to prevent a tiny increase in risk," or (ii) "deny the [transitivity] principle that, if A is better than B and B is better than C, then A must be better than C" (cf. Cibinel, 2023). These results strike many of us as highly counterintuitive.

One can also argue that fanatical theories' strange implications in choice situations like that between G and H should be blamed not on fanaticism, but rather on the unusual credence distributions of the decision makers. Perhaps it is to be expected that even the best available decision theories will produce strange implications when they are applied to unusual credence distributions. If so, then fanaticism might be a bullet worth biting.

Similar to the fanaticism problem, EBR also suffers from an apparent problem

of *excessive permissiveness*. For instance, consider a choice between two options, J and K. As before, I have 99.9% credence in T1, and 0.1% credence in T2. According to both T1 and T2, I have moderate requiring reasons in favour of J, and strong requiring reasons against K. According to T2—but not T1—I also have extremely strong justifying reasons in favour of K.[10] Numerically:

T1	Requiring strength	Justifying strength
J	10	10
K	−100	0

and

T2	Requiring strength	Justifying strength
J	10	10
K	−100	100,000

Under these assumptions, the expected strengths of the justifying and requiring reasons for my two options will be:

Expected	Requiring strength	Justifying strength
J	10	10
K	−100	100

Thus, EBR implies that J and K are both appropriate options in this choice situation, since the expected justifying strength of my reasons in favour of each option is greater than or equal to the expected requiring strength of my reasons in favour of any other option.

[10] Here is a concrete example of an extant theory of population ethics that describes the normative dimensions of one's reasons coming apart so violently. According to Thomas (2022), we have requiring reason to prevent utterly miserable and pointless lives from being caused to exist, but only purely justifying reason to create happy people. Suppose then that K involves creating many new lives, one of which isn't worth living. By contrast, we can imagine T1 as denying there is anything counting in favour of creating new happy people.

Many of us regard this implication as implausible, and intuit that option J is uniquely appropriate in this choice situation. After all, I am 99.9% sure that there are no justifying reasons in favour of K in this choice situation, and 100% certain that my requiring reasons strongly favour J over K. Under these circumstances, it seems unacceptably reckless to classify option K as appropriate.

There are (at least) three possible responses to this objection to EBR. The first two echo the two responses to the fanaticism objection that we previously mentioned (see above). Firstly, it is not clear how to avoid these results without risking other implausible results elsewhere. Secondly, the fault might lie with one's credence distribution over unusual theories, rather than with EBR.

Thirdly, EBR once again supplies additional guidance to the decision maker in this choice situation beyond the verdict that both options are appropriate. In particular, we may note that option K is *suber-appropriate* in this choice situation—insofar as in expectation I have requiring reasons against choosing K that are strong relative to my requiring reasons in favour of J. 'Suber-appropriateness' is a second-order analogue of the first-order property of *suberogation* (Atkins & Nance, 2015; Driver, 1992). Although J and K are both appropriate, there is a sense in which K is much, much worse than J. We think that this qualification makes EBR's verdict in this scenario significantly easier to swallow.

As we noted in §4 above, MEC is only applicable in cases where the decision maker has positive credence only in first-order theories like Singer's, according to which the justifying and requiring strengths of reasons are always equal to each other. EBR generalizes MEC, and it is also applicable in cases where the decision maker has positive credence in one or more first-order theories like Muñoz's according to which the justifying and requiring strengths of a reason sometimes differ.

However, this is not to say that EBR is applicable in *all* possible cases. In fact, EBR inherits from MEC an assumption that the decision maker has positive credence only in first-order moral theories between which the (requiring and justifying) strengths of reasons are *intertheoretically unit-comparable*.

Two moral theories T1 and T2 are intertheoretically unit-comparable with respect to requiring reasons iff for any options A, B, C, and D, there exists some k such that it is true and meaningful to say that the difference between the strength of reasons in favour of A and those in favour of B according to T1 is k times the size of that difference between C and D according to T2. And likewise, *mutatis mutandis*, for unit-comparability with respect to justifying reasons. Analogously, °F and °C are unit-comparable scales for temperature: a difference in temperature of 1°C is

equivalent to a difference in temperature of 1.8°F.[11]

Intertheoretic expectations are simply undefined in cases where unit comparisons are impossible. This is an important problem for MEC and EBR, since many philosophers working in the field of moral uncertainty have been sceptical about the possibility of intertheoretic unit comparisons of choiceworthiness (*inter alia*

[11]If two theories T1 and T2 are unit-comparable, then T1 and T2 must also satisfy *interval-scale measurability*. The requiring or justifying strength of reasons is interval-scale measurable according to some moral theory T iff for any options A, B, C, and D, there exists some k such that it is true and meaningful to say that the difference between the requiring or justifying strength of reasons in favour of A and those in favour of B is k times the size of that difference between C and D, according to T. Thus, temperature is one example of an interval-scale measurable variable; and °F and °C are two examples of interval scales.

These kinds of interval scales form one class of examples of *cardinal* scales. A cardinal scale represents certain properties of the subjects being measured that can only be represented by that scale and its class of *positive affine* transformations. By contrast, an *ordinal* scale represents certain properties of the subjects being measured that can only be represented by that scale and its class of *positive monotonic* transformations. The property being cardinally represented in this particular case is a reason's strength relative to other possible reasons for action.

Interval-scale measurability as defined in this footnote should not be confused with another kind of cardinal scale that economists and decision theorists are much more familiar with, *viz.* von Neumann-Morgenstern (vNM) value scales. vNM value scales represent a different property to the one represented by strength-of-reasons (choiceworthiness) interval-scales. An outcome's value on a vNM scale is a measure of how much increasing the chance of that outcome increases the value of a risky lottery. There is no guarantee that an interval-scale representation of any given moral theory will also qualify as a vNM representation—or vice versa. And there is also no guarantee that an interval-scale measurable theory will be vNM representable—or vice versa. Advocates of MEC like MacAskill et al. (2020, 7ff) are clear that they intend for MEC to operate on interval-scale representations of first-order moral theories, rather than on vNM representations.

By contrast, some other philosophers have advanced representationalist arguments for MEC. Perhaps the most sophisticated and well-developed examples of these arguments can be found in Stefan Riedener's (2021) work on uncertain values. However, as Brian Hedden (2016, p. 113) has persuasively argued, the problem with something like Riedener's approach is that it abdicates MEC's ambition "to provide a framework which takes as input an agent's credences in moral theories (and credences about descriptive matters of fact) and outputs what the agent [may appropriately] do, without presupposing any facts about what agents [may appropriately] do in various situations". What I as a morally uncertain agent want is "to be told how to start off with my credences in moral theories and use them to derive a verdict on what I [may appropriately] do, but instead the ... Riedener approach tells me that if I start off with credences in moral theories *and* facts about the "preferences" of [appropriateness], then there is a way of fixing the zero point and scale of each moral theory's value function such that [appropriateness] can be thought of as mandating [expected choiceworthiness] maximization relative to those choices of zero points and scales" (Hedden, 2016, p. 114); see also (Gustafsson, 2022, pp. 466–7).

We thank an anonymous reviewer for pressing us to clarify these points.

(Gracely, 1996; Gustafsson, 2022; Gustafsson & Torpman, 2014; Hedden, 2016)). For instance, Brian Hedden (2016) asks us to compare totalist and averagist utilitarianism. He argues that any proposed commensuration rate between total and average utility will have implausible results at certain possible sizes of the world population: "No matter what value functions we use to represent Averagism and Totalism, once we fix on some proposed decrease in average happiness, Averagism will swamp Totalism for smaller population increases while Totalism will swamp Averagism for larger population increases" (Hedden, 2016, p. 109). Thus, at least these two first-order theories appear to be intertheoretically unit-incomparable with each other. Johan Gustafsson (2022) also leans on this argument in his most recent discussion of intertheoretical incomparability.[12]

Fortunately, there are several promising responses to these kinds of objections.

Firstly, we think that the absurdities of Hedden's attempted comparisons between totalist and averagist utilitarianism are entirely attributable to the absurdities internal to averagism. To illustrate these absurdities, consider two possible worlds. In the first possible world, there is only one moral patient; in the second, they are instead one hundred billion strong. According to averagism, increasing the well-being of the moral patient in the first possible world by one util has exactly the same moral value as increasing the well-being of all hundred billion moral patients in the second possible world by one util. This *intra*theoretic unit comparison strikes many of us as absolutely absurd—despite the fact that it correctly describes the internal structure of averagism. However, if unit-comparisons between averagism's own choiceworthiness evaluations are sometimes absurd, then it should arguably come as no surprise if certain *inter*theoretic unit comparisons involving these choiceworthiness evaluations also seem absurd—even if these comparisons correctly commensurate between totalism and averagism's choiceworthiness evaluations.

Furthermore, in cases involving less absurd moral theories, intertheoretic unit-comparisons often *do* seem intuitively plausible. For instance, consider the following comparisons (all reproduced from MacAskill et al. (2020, pp. 116–7)):

> If animals have rights in the way that humans do, then killing animals is a much more severe wrongdoing than if they don't.

[12] Moreover, an anonymous reviewer points out that some moral theories apparently cannot even be represented by ordinal choiceworthiness scales—let alone be comparable intertheoretically to other moral theories. In particular, a moral theory which denies the transitivity of moral betterness cannot be represented by any numerical choiceworthiness scale. Structurally unusual moral theories like this have received little attention in the extant literature on moral uncertainty. Future work could investigate how existing decision procedures could be extended to handle these kinds of theories.

> If Singer is right about our duties to the poor, then our obligation to give to development charities is much stronger than if he's wrong.
>
> Lara used to think that stealing from big corporations was only mildly wrong, but now she thinks it outrageous.
>
> James thinks that extramarital sex is a minor wrong, but Jane thinks it's an abomination.

In each of these examples, it seems entirely run-of-the-mill to compare the strengths of moral reasons across different theories. Even if there are epistemic barriers to, for instance, knowing exactly how much stronger obligations to the poor are if Singer is correct, in these examples there do not seem to be any metaphysical problems with intertheoretic unit-comparisons.

It is still up for debate how often moral theories are intertheoretically unit-comparable with each other. However, even if this phenomenon is relatively rare, we still think that it is no small achievement to make progress on the question of which actions are appropriate for an agent who has positive credence only in unit-comparable moral theories. Making progress on this question can suggest ideas for and constraints on the project of developing a more-generally applicable theory of appropriate action (cf. Tarsney, 2021). We hope to have made that kind of progress in this paper.

In order to handle cases in which some of the moral theories in which the decision maker has positive credence are intertheoretically unit-incomparable with each other, EBR will require further modification and supplementation. Fortunately, other philosophers have already proposed modifications to MEC designed to handle these kinds of cases that could be adapted for use with EBR. For instance, one can cardinalize purely ordinal choiceworthiness orderings using the *Borda score* method (MacAskill, 2016). Also, a *variance normalization* technique can, arguably, be used to rescale intertheoretically incomparable choiceworthiness schedules so that the units of the rescaled versions are comparable across different moral theories (MacAskill et al., 2020). We hope to discuss in future work how these MEC techniques can be adapted for use with EBR.

6 Conclusion

In this paper, we have argued that decision makers who have at least some positive credence in moral theories that allow for the requiring strengths of one's reasons

to come apart from their justifying strengths should use EBR to decide which options are appropriate under moral uncertainty. We have also argued that MEC is inapplicable in cases like this, and that EBR is superior to MERS and MEW.

One important practical upshot of EBR concerns charitable activities. Imagine that "I am faced with a choice whether to spend $500 on a new TV or donate the money to GiveDirectly. I am sure that donating the money is either morally obligatory or supererogatory, but unsure which. Conversely, I am sure that buying the new TV is either morally prohibited or merely permissible, but unsure which" (Tarsney, 2019, p. 599). Some advocates of MEC—including Christian Tarsney—have argued that in this choice situation, it is uniquely appropriate for me to donate my $500 to GiveDirectly. According to Tarsney (2019, p. 599): "The argument for this claim is straightforward: because donating to charity is certainly at least as good as, and possible better than, buying the TV, the former option statewise dominates the latter"; see also Ross (2006).

Our arguments in this paper challenge Tarsney's position here. He implicitly assumes that choiceworthiness is unidimensional, and hence that the choiceworthiness of donating to GiveDirectly must be at least as great as the choiceworthiness of buying the television. However, we suggest that the supererogationist theory can instead be represented as asserting that although there are strong requiring reasons in favour of donating to charity, there are also strong justifying reasons in favour of me spending my $500 on a new TV, or the like. Under this analysis, EBR implies that it might well be appropriate under moral uncertainty for me to spend my $500 on the TV rather than on a GiveDirectly donation. Of course, donating to GiveDirectly is the super-appropriate option in this choice situation, and as such is to be encouraged. But if I do choose to spend my $500 on a new TV, then I need not be acting inappropriately in this choice situation. This strikes us as desirable result. Future work could investigate the practical implications of EBR in other areas of applied ethics.

References

Atkins, P., & Nance, I. (2015). Defending the suberogatory. *Journal of Ethics and Social Philosophy, 9(1)*, 1–7.

Barry, C., & Tomlin, P. (2016). Moral uncertainty and permissibility: Evaluating option sets. *Canadian Journal of Philosophy, 46(6)*, 898–923.

Beckstead, N., & Thomas, T. (2023). A paradox for tiny probabilities and enormous values. *Noûs*, *00*, 1–25. https://onlinelibrary.wiley.com/doi/10.1111/nous.12462.

Campbell, T., & Kaczmarek, P. (ms). Does the asymmetry condemn selfless parents? *unpublished manuscript*, *00*, 1–28.

Chang, R. (2002). The possibility of parity. *Ethics*, *112(4)*, 659–688.

Chappell, R. Y. (2015). Value receptacles. *Noûs*, *49(2)*, 322–332.

Cibinel, P. (2023). A dilemma for nicolausian discounting. *Analysis*, *00*. https://doi.org/10.1093/analys/anac095.

Dancy, J. (2004). Enticing reasons. In R. J. Wallace, P. Pettit, S. Scheffler, & M. Smith (Eds.), *Reasons and value: Themes from the moral philosophy of joseph raz* (pp. 91–118). Oxford University Press.

Driver, J. (1992). The suberogatory. *Australasian Journal of Philosophy*, *70(3)*, 286–295.

Gert, J. (2004). *Brute rationality*. Cambridge University Press.

Gracely, E. J. (1996). On the noncomparability of judgements made by different ethical theories. *Metaphilosophy*, *27(3)*, 327–332.

Greaves, H., & Ord, T. (2017). Moral uncertainty about population axiology. *Journal of Ethics and Social Philosophy*, *12(2)*, 135–167.

Gustafsson, J. (2022). Second thoughts about my favourite theory. *Pacific Philosophical Quarterly*, *103(3)*, 448–470.

Gustafsson, J., & Torpman, O. (2014). In defence of my favourite theory. *Pacific Philosophical Quarterly*, *95(2)*, 159–174.

Harman, E. (2014). The irrelevance of moral uncertainty. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics, volume 10* (pp. 53–79). Oxford University Press.

Hedden, B. (2016). Does MITE make right? In R. Shafer-Landau (Ed.), *Oxford studies in metaethics, volume 11* (pp. 102–128). Oxford University Press.

Horgan, T., & Timmons, M. (2017). Untying a knot from the inside out: Reflections on the "paradox" of supererogation. *Social Philosophy and Policy*, *27(2)*, 29–63.

Hurka, T., & Shubert, E. (2012). Permissions to do less than the best: A moving band. In M. Timmons (Ed.), *Oxford studies in normative ethics, volume 2* (pp. 1–27). Oxford University Press.

Hurley, P. (1995). Getting our options clear: A closer look at agent-centred options. *Philosophical Studies*, *78(2)*, 163–188.

Kaczmarek, P., Lloyd, H., & Plant, M. (2023). Moral uncertainty, proportionality and bargaining. *unpublished manuscript*, *00*, 1–23. https://philpapers.org/rec/KACMUP.

Kagan, S. (1989). *The limits of morality*. Oxford University Press.
Kamm, F. M. (1985). Supererogation and obligation. *Journal of Philosophy*, *82(3)*, 118–138.
Kamm, F. M. (1992). Review: Non-consequentialism, the person as an end-in-itself, and the significance of status. *Philosophy & Public Affairs*, *21(4)*, 354–389.
Lazar, S. (2013). Associative duties and the ethics of killing in war. *Journal of Practical Ethics*, *1(1)*, 3–48.
Lazar, S. (2019). Moral status and agent-centred options. *Utilitas*, *31(1)*, 83–105.
Little, M. O., & Macnamara, C. (2021). Non-requiring reasons. In R. Chang & K. Sylvan (Eds.), *The routledge handbook of practical reasons* (pp. 393–404). Routledge.
Lockhart, T. (2000). *Moral uncertainty and its consequences*. Oxford University Press.
MacAskill, W. (2016). Normative uncertainty as a voting problem. *Mind*, *125(500)*, 967–1004.
MacAskill, W. (2019). Practical ethics given moral uncertainty. *Utilitas*, *31(3)*, 231–245.
MacAskill, W., Bykvist, K., & Ord, T. (2020). *Moral uncertainty*. Oxford University Press.
MacAskill, W., & Ord, T. (2020). Why maximize expected choice-worthiness? *Noûs*, *54(2)*, 327–353.
Mogensen, A. (2019). Staking our future: Deontic long-termism and the non-identity problem. *Global Priorities Institute Working Paper No. 9-2019*, *00*, 1–32. https://globalprioritiesinstitute.org/andreas-mogensen-staking-our-future-deontic-long-termism-and-the-non-identity-problem/.
Muñoz, D. (2021). From rights to prerogatives. *Philosophy and Phenomenological Research*, *102(3)*, 608–623.
Muñoz, D., & Pummer, T. (2022). Supererogation and conditional obligation. *Philosophical Studies*, *179(5)*, 1429–1443.
Muñoz, D., & Spencer, J. (2021). Knowledge of objective 'oughts': Monotonicity and the new miners puzzle. *Philosophy and Phenomenological Research*, *103(1)*, 77–91.
Parfit, D. (1988). What we together do. *unpublished manuscript*, *00*, 1–34. https://philarchive.org/archive/PARWWT-3.
Parfit, D. (2011). *On what matters, volume 1*. Oxford University Press.
Podgorski, A. (2020). Normative uncertainty and the dependence problem. *Mind*, *129(513)*, 43–70.
Pummer, T. (2023). *The rules of rescue: Cost, distance, and effective altruism*. Oxford University Press.

Regan, D. H. (1980). *Utilitarianism and co-operation*. Oxford University Press.
Riedener, S. (2021). *Uncertain values: An axiomatic approach to axiological uncertainty*. de Gruyter.
Ross, J. (2006). Rejecting ethical deflationism. *Ethics, 116(4)*, 742–768.
Russell, J. S. (2023). On two arguments for fanaticism. *Noûs, 00*, 1–31. https://doi.org/10.1111/nous.12461.
Scheffler, S. (1982). *The rejection of consequentialism: A philosophical investigation of the considerations underlying rival moral conceptions*. Oxford University Press.
Schofield, P. (2021). *Duty to self: Moral, political, and legal self-relation*. Oxford University Press.
Sepielli, A. (2010). *Along an imperfectly lighted path: Practical rationality and normative uncertainty* [Doctoral dissertation, Rutgers University].
Slote, M. (1984). Morality and self-other asymmetry. *Journal of Philosophy, 81(4)*, 179–192.
Sung, L. (ms). Against maximize expected choiceworthiness. *unpublished manuscript, 00*, 1–36.
Tarsney, C. (2019). Rejecting supererogationsim. *Pacific Philosophical Quarterly, 100(2)*, 599–623.
Tarsney, C. (2021). Vive la différence? structural diversity as a challenge for metanormative theories. *Ethics, 131(2)*, 151–182.
Thomas, T. (2022). The asymmetry, uncertainty and the long term. *Philosophy and Phenomenological Research, 107(2)*, 470–500.
Weatherson, B. (2019). *Normative externalism*. Oxford University Press.
Wilkinson, H. (2022). In defense of fanaticism. *Ethics, 132(2)*, 445–477.
Wilkinson, H. (2023). Egyptology and fanaticism. *Global Priorities Institute Working Paper No. 12-2023, 00*, 1–23. https://globalprioritiesinstitute.org/egyptology-and-fanaticism-hayden-wilkinson/.