Effective Altruism
and
Decision-Making for the Clueless

Dimitri J. van Capelleveen

Supervisor: Dr. Conrad Heilmann

Advisor: Dr. Stefan Wintein

Third Reader: Dr. Christopher Clarke

30,000 words

July 8th, 2022

Abstract. Various philosophers have recently claimed that normative uncertainty in general, or decision-theoretic, moral, or prudential uncertainty in particular, should be taken into consideration in decision-making, a position sometimes labelled Metanormativism. In this thesis, I argue against a restricted form of Metanormativism, pertaining only to uncertainty about imprecise decision principles, i.e., accounts of rational choice for decisions modelled using imprecise credences ('decisions under cluelessness'). To factor uncertainty about imprecise decision principles into decision-making, a Metanormativist account of rational choice is required, which tells decision-makers how to act on the basis of (among other things) uncertainty about imprecise decision principles. Such a Metanormativist account of rational choice can take one of two general forms. First, it might take the form of a single second-order decision principle, i.e., an account that specifies which alternatives are permissible on the basis of our confidence in imprecise decision principles and the permissibility of alternatives according to these principles. Second, the Metanormativist account could be comprised of an infinite hierarchy of second- and higher order decision principles: In case decision-makers are uncertain about second-order decision principles, decision-makers should resort to a third-order decision principle; uncertainty about third-order decision principles should be handled by means of a fourth-order decision principle; and so on. I argue that there isn't any single second-order decision principle that decision-makers should conform to, nor any plausible account of choice comprised of an infinite hierarchy of second- and higher order decision principles. Hence there is no plausible Metanormativist account of rational choice, from which it follows that Metanormativism about imprecise decision principles is false. To arrive at this conclusion, I'll have to lay considerable groundwork, since the possibility of imprecise credences has so far only been acknowledged in passing in the literature on Metanormativism. Notably, I don't hold the discussion about Metanormativism in the abstract, but embed it in the context of effective altruism. Considering that effective altruists often face decisions best modelled using imprecise credences, the conclusion that Metanormativism about imprecise decision principles is false has important implications for this social movement.

Acknowledgments. I'm very grateful to Conrad Heilmann for comments on many previous drafts and many helpful discussions; to Stefan Wintein for comments on a previous draft; to Akshath Jitendranath for comments on part of a previous draft and discussion of its content; and to my parents for their continuous and loving support.

§1. Introduction

Effective altruism is both an intellectual and a practical project (e.g., MacAskill 2018, 442; 2019, 12*ff*.; MacAskill and Pummer 2020; cf. Singer 2015, 4–5; Berkey 2021). As an intellectual project, or research field, effective altruism uses evidence and reasoning to determine how to do the most good. Typically, it's assumed that the good (i.e., value) is constituted by well-being and that everyone's well-being is of equal moral importance. As a practical project, or social movement, effective altruism puts the lessons drawn by its intellectual counterpart into practice.

At the time of writing, about 7,400 people reportedly self-identify as 'effective altruists,' and an estimated $46 billion is in one way or another committed to effective altruism (Todd 2021b). So, clearly, effective altruism is booming–it has the potential to have a major impact through its practical branch. Hence, it's worth investigating its underpinnings on the intellectual side.

Many of those engaged in effective altruism as an intellectual project rely on decision theory as a foundation for their views (MacAskill and Pummer 2020, 4). As Sebo and Paul (2019, 53) and Gabriel and McElwee (2019, 100) have pointed out, just like (orthodox) decision theorists, many effective altruists endorse

> *Maximize Expected Value (informal).* An alternative x is rationally permissible iff x maximizes expected value.

It should therefore be unsurprising that in research conducted by effective altruists, Maximize Expected Value (MEV) is commonly assumed (see, e.g., Tomasik 2014; Askell 2019, 38;

Snowden 2019, 70*ff.*; Bruers 2019; Tarsney 2020, 4; MacAskill and Greaves 2021, 5; see also Karnofsky 2016; GiveWell 2017; Todd 2021a; 2021b).

Recently, though, Greaves (2016, 323*ff.*) has argued that when effective altruists come to their decisions, MEV is (often) inapplicable, using argumentation subsequently echoed by Herlitz (2019, 5–9)[1] and Mogensen (2021, 143–146). Greaves argues that effective altruists (often) face decisions under *cluelessness*, sometimes also called decisions under *deep* or *severe uncertainty.* Key is that these decisions are best modelled using so-called *imprecise credences.* Importantly, if imprecise credences are assigned, the expected value of alternatives isn't well-defined. It follows that MEV is inapplicable.

In this thesis–after delving into the relationship between effective altruism and decision theory, explaining the notion of imprecise credence, and providing a more precise characterization of MEV–, I'll take off from the claim that effective altruists (often) face decisions under cluelessness. This means that effective altruists are in need of some decision principle to replace MEV. Decision theorists have proposed numerous–generally quite complicated–*imprecise decision principles*, each of which specifies how one should rationally act if confronted with decisions modelled using imprecise credences (see, e.g., Elga 2010). Greaves (2016, 329; 333–334) tentatively endorses one ('Moderate'), while Herlitz (2019, 13–14) and Mogensen (2021, 146–151) tentatively endorse another ('Maximality'). It stands to reason that one or other imprecise decision principle should replace MEV. It's unclear, however, which it should be (Moderate? Maximality?). If this isn't due to irrationality on our part, it must be possible to be *rationally uncertain* about which of these principles to conform to (§2).

This means that it can be rational for effective altruists to be *normatively uncertain*, i.e., uncertain about *what they ought to do*, or, more narrowly, *decision-theoretically uncertain*, i.e., uncertain about *what they ought rationally to do.* There has been increasing interest in decision-making under normative or decision-theoretic uncertainty. One might hold that such uncertainty should be accounted for in decision-making, a position variously referred to as *Metanormativism*, *Uncertaintism*, or *Normative Internalism.* Views of this sort have been endorsed by, among others, Robert Nozick, William MacAskill, Christian Tarsney, Andrew Sepielli, Hillary Greaves, Toby Ord, Kirster Bykvist, Abelard Podgorski, Philip Trammell, Stefan Riedener, Caspar Oesterheld, Aaron Vallinder, Carl Shulman, Johannes Treutlein, and

[1] The page numbers for Herlitz (2019) refer to the preprint.

Andreas Mogensen (see Nozick 1995, 34–50; MacAskill 2014; 2016b; Tarsney 2019a; Sepielli 2014; Greaves and Ord 2017; MacAskill and Ord 2020; MacAskill, Bykvist, and Ord 2020; Podgorski 2020; Trammell 2021; Riedener 2021; MacAskill *et al.* 2021; Mogensen 2021, fn. 16).

There has also been a backlash against this whole project; that is, some philosophers have denied that normative or decision-theoretic uncertainty should be accounted for in decision-making, a position sometimes called *Actualism* or *Normative Externalism*. Views of this sort have been taken up by philosophers such as Brian Weatherson, Elizabeth Harman, and Brian Hedden (see Weatherson 2014; 2019; Harman 2015; Hedden 2016).

More to the point, one might endorse or, alternatively, deny

> *Metanormativism about Imprecise Decision Principles.* Rational uncertainty about imprecise decision principles should rationally be accounted for in decision-making.

With regards to effective altruism, the decision of whether to accept or reject Metanormativism about Imprecise Decision Principles (Metanormativism for short) has potentially immense implications. If effective altruists were to include rational uncertainty about imprecise decision principles in their decision-making, they might end up making wildly different choices than if they were to exclude it. As it happens, Metanormativism, or positions similar to it, have been publicly endorsed by various prominent and influential members of the effective altruism community, most notably MacAskill, Tarsney, Ord, Greaves, and Mogensen (§3).

Nevertheless, I'll argue against Metanormativism in this thesis. To show that Metanormativism is false, I'll argue that there's no plausible Metanormativist account of rational decision-making, i.e., no plausible account that tells us what to do on the basis of (among other things) rational uncertainty about imprecise decision principles. There are two important, general forms that such a Metanormativist account might take, and I'll criticize both options in turn.

First, it might take the form of a single *second-order decision principle*, sometimes also called a *meta decision theory.* Roughly, a second-order decision principle is an account that specifies which alternatives are rationally permissible on the basis of (i) our confidence in imprecise decision principles and (ii) the permissibility of the available alternatives according to these imprecise decision principles. An example of a second-order decision principle is

> *My Favourite Theory.* An alternative *x* is rationally permissible iff *x* is rationally permissible according to the (an) imprecise decision principle that the decision-maker rationally finds most plausible.

Tarsney (2019) has basically argued that there's a unique second-order decision principle that decision-makers should conform to. Call such an account a *Unique Account.* I'll argue that there isn't any second-order decision principle that decision-makers should under all circumstances conform to in order to account for their uncertainty about imprecise decision principles, i.e., that there's no plausible Unique Account.

Notably, discussion about decision-making under normative or decision-theoretic uncertainty has proceeded on the assumption that decision-makers assign *precise* rather than *imprecise* credences. To my knowledge, the possibility of imprecise credences has so far only been acknowledged in passing (see, e.g., MacAskill and Ord 2020, 329; MacAskill 2016a, fn. 52; see also Mogensen 2021, fn. 16). So, to arrive at the rejection of Unique Accounts, I'll have to lay some groundwork. In particular, I'll have to develop some second-order decision principles that can accommodate imprecise credences.

For this purpose, I'll first focus on a cousin of decision-theoretic uncertainty, *moral uncertainty*, i.e., uncertainty about *what ought morally to be done.* My Favourite Theory (MFT) has, as applied to decision-making under moral uncertainty, been under severe criticism. I survey the literature and show for three objections that each has been used to support an alternative second-order decision principle, thereby gathering a small collection of such principles. (As an aside, it's noteworthy that Mogensen (2021) spend considerable space exploring the practical implications of the presence of cluelessness on effective altruist decision-making. In doing so, Mogensen implicitly relied on MFT. The objections raised against MFT prove this view to be inadequate, which undermines much of Mogensen's discussion pertaining to effective altruist decision-making under cluelessness). Importantly, though, the resulting second-order decision principles will only be able to deal with precise rather than imprecise credences. But, I'll show that they can handle the imprecise credence case once they are combined with imprecise decision principles. So, each proposed second-order decision principle will essentially be *a variation on an imprecise decision principle.* While Metanormativism claims that rational uncertainty about imprecise decision principles should be accounted for in decision-making, Unique Accounts must therefore maintain that rational uncertainty about second-order decision principles inspired *by these very same imprecise decision principles shouldn't* *be accounted for in decision-making*. This is an unpalatable implication, which should (among other reasons) lay to rest the claim that there is some suitable Unique Account (§4).

The second general form that the Metanormativist account of decision-making might take is that of *an infinite hierarchy of second- and higher order decision principles* that decision-makers have confidence in. Call such an account a *Hierarchy Account.* Positions of this sort have been taken by Sepielli (2014), MacAskill (2014, 217–219), and Trammell (2021). To see what I have in mind, note that in case decision-makers are uncertain about second-order decision principles, the Metanormativist might admit that a single second-order decision principle cannot do the trick. She might claim that, instead, the decision-maker should resort to a *third-order decision principle* to come to a decision. Roughly, a third-order decision principle is an account that specifies which alternatives are rationally permissible on the basis of (i) confidence in *second-order* decision principles and (ii) the permissibility of the available alternatives according to these second-order decision principles. Admitting that decision-makers may need to make use of a third-order decision principle paves the way for *infinite regress*: Uncertainty about third-order decision principles should be handled by means of a fourth-order decision principle, and so on. Nonetheless, the Metanormativist might hold that, in the face of such infinite regress, rational decision-making isn't (necessarily) impaired.

I'll show, however, that the infinite regress is *vicious*, i.e., that there cannot be a plausible Hierarchy Account. I'll focus on one particular case in which infinite regress arises. In this case, the decision principles will disagree about which alternatives are rationally permissible at all orders. As a result, it's left unspecified whether any alternative is rationally permissible. This violates

> *Decisiveness (informal)*. In any decision-situation, at least one available alternative is permissible.

However, the case under consideration is such that the decision-maker is rationally *only* confident in decision principles that meet Decisiveness, at every order. So, this decision-maker can be rationally certain that Decisiveness is true. Nevertheless, if what she should rationally do is responsive to her uncertainty about decision principles–in her case, a state of rational uncertainty that entails that Decisiveness must be true–, then Decisiveness must be false. This implication is unacceptable, or so I'll argue. Hence, there's no plausible Hierarchy Account (§5).

Given that both general forms that the Metanormativist account of decision-making might take (i.e., Unique Accounts, Hierarchy Accounts) ought to be rejected, so should Metanormativism. To wrap up, I'll–tentatively–tease out some implications of this conclusion for effective altruism. Most importantly, I'll take it that the failure of Metanormativism doesn't entail that cluelessness (i.e., uncertainty about outcomes best modelled using imprecise credences) is irrelevant for decision-making. It seems that decisions under cluelessness should be made in accordance with an adequate (true) imprecise decision principle, even if it's utterly unclear what this principle or its prescriptions might be. Consequently, the failure of Metanormativism entails that effective altruists are often *unavoidably clueless* in their decision-making. This conclusion will be hard to swallow for many effective altruists, so to accommodate them, I'll close by tentatively suggesting a possible escape route (§6).

§2. Preliminary

§2.1 Decision Theory for Effective Altruists

Before getting into the core arguments against Metanormativism, and before even precisely formulating the possible Metanormativist accounts of decision-making (i.e., Unique Accounts, Hierarchy Accounts), we'll need to do considerable preliminary work. For starters, it'll be helpful to delve into the relationship between effective altruism and decision theory: Why would effective altruists need a decision principle, such as MEV, in the first place? This is basically the question that I'll address in the present subsection. Next, to state MEV precisely, I'll need to explain the concept of *credence*. I'll tackle this in §2.2. I'll then give a precise characterization of MEV in §2.3, and proceed on the assumption that MEV is typically inapplicable when effective altruists have to come to decisions. As we'll see, an obvious substitute for MEV is one or other *imprecise decision principle*. In §2.4, I'll develop a formal framework with which we can state imprecise decision principles precisely. In §2.5, I'll give some concrete examples of imprecise decision principles. Regrettably, it's unclear which imprecise decision principle to employ, seemingly making it possible to be rationally uncertain about which imprecise decision principle to conform to in the face of decisions under cluelessness. This raises the question of whether such uncertainty should be accounted for in decision-making, i.e., whether Metanormativism is true; a question that I'll tackle from the next section onwards. So, let's get to it.

Emerging in large part due to Peter Singer and his 1972-article 'Famine, Affluence, and Morality' and the additional efforts of, among others, Will MacAskill and Toby Ord (for more details, see, e.g., Centre for Effective Altruism n.d.; Singer 2015, Ch. 2; Lichtenberg 2015), effective altruism has been embraced by many people, including many academics. Effective altruism is, both as an intellectual project and as a practical project, quite pluralistic: Effective altruists concern themselves with lots of different things, take lots of different approaches, and have lots of different world views. Hence making sense of the value of decision theory for effective altruists will involve some generalizations. I've attempted to formulate my claims so that they won't offend (too many) effective altruists. My way of viewing things is at least a reasonable interpretation of how effective altruists tend to operate.

Now, effective altruists involved in the intellectual project will, by definition, use 'evidence and reasoning to determine how to do the most good.' Otherwise put, effective altruists involved in the intellectual project will use evidence and reasoning for the sake of *action guidance*, i.e., to help them identify which actions to perform, in order to do the most good. While these actions should be morally right, effective altruism isn't just a moral undertaking. Many effective altruists care deeply about *rationality*. That is, effective altruists typically want to act rationally as well morally. Such effective altruists, when involved in the intellectual project, will use evidence and reasoning for the sake of moral and rational action guidance, i.e., to help them identify which actions to perform, in order to do the most good, both morally and rationally speaking. I'll be concerned with rationality. It's beyond the scope of this thesis to delve into the meaning of the notion of 'rationality,' so suffice it to say that actions are commonly perceived as rational (irrational) iff, roughly, they (don't) *cohere* with our beliefs about our aims. For instance, an effective altruist who is certain that making a donation to the Against Malaria Foundation will do the most good is rational (irrational) just in case she does (doesn't) make the donation. In addition, I'll understand 'rational' as *overall* or *all-things-considered* or, better yet, all-*rationally-relevant*-things-considered rational. So, roughly, we can say that an action is rational (irrational) iff it coheres (doesn't cohere) with all our–relevant–beliefs about our aims.

So, then, how should effective altruists go about identifying the rational actions that they should perform? For this purpose, effective altruists need some sort of decision principle. In particular, they require a decision principle that specifies which alternatives are rationally permissible, and which rationally impermissible. Decision theory can be of help here. The question of whether performing certain actions is or isn't rational falls squarely within its domain. The main take away from decision theory for effective altruists has been the importance of the decision principle that I've referred to as MEV. To state it more precisely, I'll first explain the much used notion of *credence*.

§2.2 Credences

Credences–sometimes also called *degrees of belief*, *partial beliefs*, or *subjective probabilities*–are propositional attitudes, similar to other such attitudes like belief or knowledge. So, just as you can belief or know the proposition that it's raining outside, you can give credence to this proposition. Unlike belief and knowledge, however, credence isn't simply binary. That is, while you either belief (or know) a proposition or you don't, you can give more or less credence to propositions. More precisely, credences can be expressed numerically, on a scale ranging from 0 to 1. So, you could give credence 0.6 to the proposition that it's raining outside, or credence 0.789. The higher the numerical value, the greater the confidence that a proposition is true; while credence 0 indicates that you are certain of its falsehood, credence 1 indicates full certainty of its truth. It's also possible not to give credence to propositions, i.e., to withhold judgment (just as you don't need to belief or disbelief every proposition). If you assign credence greater than 0 to a proposition, you are sometimes said to assign it *positive* or *nonzero* credence. For brevity, I'll use the term 'credence' as shorthand for 'positive credence' or 'nonzero credence,' unless explicitly stated otherwise.

Our credences towards propositions can be captured by so-called *credence functions*. These are essentially the same as probability functions; so, a credence function is a function C from a set of propositions to $\mathbb{R}$ in the interval $[0, 1]$ that satisfies the Kolmogorov probability axioms.[2]

Notably, Kolmogorov's probability axioms make demands about what our credences should rationally be like. The available evidence about the truth-status of propositions affects what credences we can rationally give to them (credences should rationally *cohere* with the available evidence; cf. the rough definition of 'rational actions' from the previous subsection). For example, if we can see that it rains outside, it would be irrational to give credence 0 to the proposition that it doesn't.

Occasionally, the available evidence warrants (only) *precise credences*, sometimes also called *sharp credences*. Credence towards a proposition is precise iff it can be represented using a single credence function. If I'm about to flip a coin that you know not to be biased, then your credences towards the proposition that it will land heads should rationally be precise: The proposition should be given credence 0.5. A credence function that can be used to represent your propositional attitudes should rationally contain only this credence towards the proposition that the coin will land heads.

Occasionally, the available evidence warrants (only) *imprecise credences*, sometimes also called indeterminate credences, unsharp credences, or mushy credences. Credence towards a propo-

[2] The Kolmogorov probability axioms are as follows (note that I mention them for completeness, not because I'll make explicit use of them): *Non-Negativity*, i.e., for any proposition P, $C(P) \geq 0$; *Normalization*, i.e., roughly, for any tautology T, $C(T) = 1$; *Finite Additivity*, i.e., for any mutually exclusive propositions P and Q, $C(P \text{ or } Q) = C(P) + C(Q)$.

sition is imprecise iff it can only be represented using a *representor*, sometimes also called a credal set, i.e., a non-empty set of credence functions C_1, C_2, …, C_n.[3] To borrow an amusing example from Elga, suppose that

> A stranger approaches you on the street and starts pulling out objects from a bag. The first three objects he pulls out are a regular-sized tube of toothpaste, a live jellyfish, and a travel-sized tube of toothpaste. To what degree should you believe that the next object he pulls out will be another tube of toothpaste? The answer is not clear. The contents of the bag are clearly bizarre. You have no theory of "what insane people on the street are likely to carry in their bags," nor have you encountered any particularly relevant statistics about this. The situation doesn't have any obvious symmetries, so principles of indifference seem to be of no help. (Elga 2010, 1)

At least intuitively, it seems that in this case, you aren't warranted to give precise credence to the proposition that the next object pulled from the stranger's bag will be another tube of toothpaste. Instead, it seems that rationality requires you to assign imprecise credences. For example, your representor might contain credence functions that assign it anything between credence 0.2 and 0.95.

In this thesis, I'll be concerned exclusively with *rational* credences. For brevity, when I'll claim that credences are or should be assigned, I'll typically omit the qualification that credences are rationally assigned or should rationally be assigned.

It's possible to assign credences to *outcomes* that we might bring about by choosing the alternatives available to us. (So, to be sure, I'm saying here that it's possible to *rationally* assign credences to outcomes). For example, I might assign credence 0.9 to the outcome that I'll be hungry in the evening given that I choose not to eat anything in the afternoon. (It's more precise to say that I might assign credence 0.9 to the *proposition* that I will be hungry in the evening conditional on the proposition that I don't eat anything in the afternoon, but it's more straightforward to talk about credences being assigned to outcomes given that we choose alternatives). With this in hand, we can precisely define MEV.

[3] Imprecise credences are sometimes formalized differently, but this isn't important for present purposes.

§2.3 Maximize Expected Value, Precisely Put

Informally, an alternative's *expected value* is equal to the sum of, for all outcomes that choosing the alternative might bring about, (i) the credence that the decision-maker assigns to the outcome given that she'll choose the alternative multiplied by (ii) the value of the outcome. Formally, the expected value EV of an alternative x is

$$\mathrm{EV}(x) = \sum_{i=1}^{n} \mathrm{C}(O_i|x)\mathbf{V}(O_i)$$

where n is the number of outcomes to which x might lead, $\mathrm{C}(\mathrm{O}_i|x)$ is the credence the decision-maker gives to outcome O_i conditional on choosing x, and $\mathbf{V}(\mathrm{O}_i)$ is the value of O_i. $\mathbf{V}$ is a *value function*, i.e., a function from outcomes to $\mathbb{R}$, where the greater the numerical value assigned to $\mathbf{V}(\mathrm{O}_i)$, the greater the value (i.e., moral value) of O_i. More specifically, I'll assume that value is *interval-scale measurable*, i.e., that a value function $\mathbf{V}$ can be transformed into a value function $\mathbf{V}'$ by means of a positive linear transformation, i.e., iff $\mathbf{V}'=a\mathbf{V}+b$, where $a>0$ (see Resnik 1987, 30). Precisely put, then, the decision principle most popular among effective altruists is

> *Maximize Expected Value (formal).* An alternative x is permissible iff there's no available alternative y such that $\mathrm{EV}(y)>\mathrm{EV}(x)$.

Two things should be noted about MEV. First, given our interest in the all-relevant-things-considered rationality of actions, the term 'permissible' should be interpreted as 'all-rationally-relevant-things-considered rationally permissible.' (This applies to my usage of the term permissible, and impermissible, throughout the thesis.) So, according to MEV, the only 'rationally relevant things' are the available alternatives, the outcomes to which they might lead, the values of these outcomes, and the precise credences given to outcomes conditional on choosing the alternatives that might bring them about.

Second, the expected value of any alternative is, by definition, established relative to a single credence function. The means that, firstly, if precise credences are assigned to all outcomes conditional on selecting the alternatives that might bring them about (and if there's a value function), the expected value of each alternative is well-defined. Hence MEV is applicable. In contrast, it doesn't always make sense to talk about "an alternative's expected value" in case imprecise credences are assigned. Different credence functions in a decision-maker's representor can disagree about what the alternative's expected value is. So, instead of having a single expected value assigned to the alternative, we can end up with a set of multiple, distinct expected values, corresponding to the different credence functions in the representor. But MEV isn't designed to deal with multiple, distinct expected values for the same alternative and, therefore, in the imprecise credence case, MEV isn't applicable (see S. Bradley and Steele 2014, 278).

Greaves (2016, 323*ff.*) has argued that when effective altruists come to their decisions, MEV is (often) inapplicable, with Herlitz (2019, 5–9) and Mogensen (2021, 143–146) following in her footsteps. The reason for the (frequent) inapplicability of MEV is that when effective altruists come to their decisions, their decisions are (often) best modelled using imprecise credences. In line with Greaves, Herlitz, and Mogensen, I'll refer to these as *decisions under cluelessness.*

Now, Greaves isn't very explicit on how ubiquitous decisions under cluelessness are for effective altruists, though she gives the impression that, in her view, they have to be dealt with quite often. For instance, Greaves says that "cases with the structure in question [i.e., where decisions under cluelessness have to be made] also occur in myriad other decision contexts [i.e., outside of effective altruism], at both large and small scales. For example: […] An individual's decision as to which degree course to sign up for, which job to accept, whether or not to have children, how much to spend on clothes, whether or not to give up caffeine" (334). Herlitz (2019) is more explicit, writing that "cluelessness […] typically appears in the areas that effective altruists tend to be most interested in: global health, development, the future of humanity, research and development, existential risk, climate change" (9). Mogensen (2021) concludes for various real-life charitable organizations that, roughly, if effective altruists were to choose between them, they would face decisions under cluelessness.

Whether or not effective altruists are indeed often confronted with decisions under cluelessness isn't of the utmost importance (although the stakes would be considerably higher if it were so). What matters is that for these cases, effective altruists are in need of some decision principle to replace MEV for the purpose of rational action guidance. This is where *imprecise decision principles* come in.

§2.4 A Formal Framework for Imprecise Decision Principles

There's an ongoing debate among decision theorists about how to come to rational decisions when imprecise credences are assigned to outcomes, i.e., about the appropriate imprecise decision principle (rule, criterion, theory). Recent discussion was sparked by an influential paper by Elga (2010), who argued that there's no plausible account for rational decision-making using imprecise credences. In response, various accounts have recently been defended by, among others, Rinard (2015). It's reasonable to suppose that such a principle should substitute MEV when effective altruists face decisions under cluelessness.

So, in this subsection and the next, I'll give a brief survey of the literature on imprecise decision principles. In this subsection, I'll introduce some formal terminology that we'll use to state the principles in a precise manner. The literature conventionally focuses on imprecise decision principles that assess alternatives on the basis of their expected values relative to some (not necessarily all) credence functions in the representor, for reasons that will become apparent shortly. I'll stick to this convention and, therefore, have developed a formal framework that gives us the tools to define this type of principle. In the next subsection, I'll give some concrete examples of imprecise decision principles. I'll concentrate on the imprecise decision principles–tentatively–adopted by Greaves (2016, 329; 333–334), Herlitz (2019, 13–14), and Mogensen (2021, 146–151), since they share my interest in effective altruist decision-making under cluelessness. (For additional examples of imprecise decision principles, see, e.g., Weatherson 1998; Elga 2010; S. Bradley and Steele 2014; Rinard 2015; Greaves 2016, 328–29; R. C. Bradley 2017, 271–77; Mogensen 2021, 146–51).

Here's the formal framework that I've developed.[4] Let $\mathfrak{O}$ be the set of all possible outcomes (states of affairs, possible worlds) O_1, O_2, …, O_n. Let $\mathfrak{A}$ be the set of all possible alternatives (actions, options) x, y, z, etc., such that for each alternative $x \in \mathfrak{A}$, a non-empty subset of $\mathfrak{O}$ contains the mutually exclusive and jointly exhaustive outcomes that x might bring about. Informally, let a *decision-situation* be any case in which a decision-maker has to come to a decision under cluelessness. Formally, let a decision-situation $\mathfrak{D}$ be a tuple $(\mathbf{D}, \mathbf{A}, \mathbf{O}, \mathbf{R}, \mathbf{V})$, where $\mathbf{D}$ is a decision-maker; $\mathbf{A}$ is a set of available alternatives, i.e., a subset of $\mathfrak{A}$, such that $|\mathbf{A}| \geq 2$; $\mathbf{O}$ is a subset of $\mathfrak{O}$, such that it contains all and only the outcomes that might be

[4] The framework is, of course, not wholly original. In particular, I've drawn on MacAskill (2016a, 969), who formalizes the notion of a 'decision-situation' in a similar manner, albeit for a different purpose.

brought about by the alternatives in $\mathbf{A}$; so, for each alternative $x \in \mathbf{A}$, a subset of $\mathbf{O}$ contains the mutually exclusive and jointly exhaustive outcomes that x might bring about; $\mathbf{R}$ is a subset of $\mathbf{D}$'s representor, such that $\mathbf{R}$ contains only the credences that imprecise decision principles might deem relevant for coming to a decision; that is, each credence function $C_i \in \mathbf{R}$ is such that it assigns only rationally admissible numerical values to, for each $x \in \mathbf{A}$ and for each $O_i \in \mathbf{O}$ that might be brought about by x, $C_i(O_i|x)$; $\mathbf{V}$ is as before (see §2.3, or, for your convenience, this footnote)[5] with the addition that it's tailored to this particular decision-situation, in the sense that it contains only the values that imprecise decision principles might deem relevant for coming to a decision; that is, $\mathbf{V}$ is such that for each $O_i \in \mathbf{O}$, $\mathbf{V}(O_i)$ is assigned a rationally admissible numerical value.

To illustrate the notion of 'decision-situation,' consider the following concrete case (which I'll use as a running example). Suppose that an effective altruist, Nina (the decision-maker $\mathbf{D}$), has to decide between making a €100 donation to either the Against Malaria Foundation (alternative x), Animal Charity Evaluators (alternative y), or the Patient Philanthropy Fund (alternative z) (in which case $\mathbf{A}=\{x, y, z\}$ and $|\mathbf{A}|=3$). As a result of x, the Against Malaria Foundation would use Nina's donation to purchase and distribute anti-malarial bed nets, thereby either saving one child from death by malaria, such that this child would go on to have two children, three grandchildren, and four great-grandchildren (outcome O_1) or saving two children from death by malaria, such that these children would go on to have a total of four children, six grandchildren, and eight great-grandchildren (outcome O_2). As a result of y, Animal Charity Evaluators would conduct research that either leads to the prevention of the birth of ten nonhuman animals, who would have lived and died in factory farms (outcome O_3) or twenty nonhuman animals, who would have suffered the same fate (outcome O_4). Finally, as a result of z, the Patient Philanthropy Fund would invest Nina's donation for three hundred years, after which they'd use it to support research into moral philosophy, that would either make the grantees a little happier (outcome O_5) or a lot happier (outcome O_6) (hence $\mathbf{O}=\{O_1, O_2, O_3, O_4, O_5, O_6\}$). Nina's representor is such that it contains two relevant credence functions (i.e., relevant as far as imprecise decision principles are concerned). Credence function C_1 is such that $C_1(O_1|x)=0.5$; $C_1(O_2|x)=0.5$; $C_1(O_3|y)=0.7$; $C_1(O_4|y)=0.3$; $C_1(O_5|z)=0.1$;

[5] $\mathbf{V}$ is a value function, i.e., a function from outcomes to $\mathbb{R}$, where the greater the numerical value assigned to $\mathbf{V}(O_i)$, the greater the value (i.e., moral value) of O_i, such that that a value function $\mathbf{V}$ can be transformed into a value function $\mathbf{V}'$ by means of a positive linear transformation.

$C_1(O_6|z)=0.9$. Credence function C_2 is such that $C_2(O_1|x)= 0.4$; $C_2(O_2|x)=0.6$; $C_2(O_3|y)= 0.8$; $C_2(O_4|y)=0.2$; $C_2(O_5|z)=0.1$; $C_2(O_6|z)=0.9$ (so, $\mathbf{R}=\{C_1, C_2\}$). Finally, $\mathbf{V}$ is such that $\mathbf{V}(O_1)=5$; $\mathbf{V}(O_2)=10$; $\mathbf{V}(O_3)=6$; $\mathbf{V}(O_4)=12$; $\mathbf{V}(O_5)=4$; $\mathbf{V}(O_6)=8$ (or any positive linear transformation thereof).

On these assumptions, each alternative has two expected values, one relative to C_1, and another relative to C_2. Relative to C_1, the expected value of y is 7.8, that of z is 7.6, and that of y is 7.5. Relative to C_2, the expected value of x is 8, that of z is 7.6, and that of y is 7.2. Hence y maximizes expected value relative to C_1 and x relative to C_2, and z relative to neither, though it has greater expected value than x and y relative to the credence functions according to which these alternatives don't maximize expected value.

Now, as a first approximation, let an *imprecise decision principle* P^1 be an account that specifies, for every possible $\mathfrak{D}$, which alternatives are permissible and impermissible for $\mathbf{D} \in \mathfrak{D}$. The permissible alternatives according to an imprecise decision principle P^1 go into what we'll call its *choice-set* CS_{P1}, i.e., a subset of $\mathbf{A} \in \mathfrak{D}$ associated with P^1. In addition, the impermissible alternatives according to an imprecise decision principle P^1 go into what we'll call its *prohibited-set* PS_{P1}, also a subset of $\mathbf{A} \in \mathfrak{D}$ that is associated with P^1.[6] Since the same alternative x in the same decision-situation $\mathfrak{D}$ cannot be simultaneously declared permissible and impermissible by the same imprecise decision principle P^1, let CS_{P1} and PS_{P1} be such that if $x \in PS_{P1}$, then $x \notin CS_{P1}$, and if $x \in CS_{P1}$, then $x \notin PS_{P1}$.

To illustrate, any imprecise decision principle P^1 should specify in Nina's decision-situation (and in any other) which alternatives are permissible and impermissible. If P^1 declares only x permissible, then $CS_{P1}=\{x\}$. Hence it cannot be impermissible, i.e., an element of PS_{P1}. If P^1 says that y and z are impermissible, then $PS_{P1}=\{y, z\}$, meaning neither can be permissible, i.e., elements of CS_{P1}.

One might object that the concept of a prohibited-set is redundant, since if an alternative isn't permissible, it must be impermissible, and *vice versa.* This is why MEV specified only when alternatives are permissible; it was implicit that alternatives that aren't permissible, are impermissible. As we'll see shortly, however, not all imprecise decision principles behave like this, and hence the concept of a prohibited-set is a useful part of our toolkit.

[6] The notion of a 'prohibited-set' is borrowed from Tarsney (2019, 4).

Recall from §2.3 that the terms 'permissible' and 'impermissible' should be understood as 'all-rationally-relevant-things-considered rationally permissible' and 'all-rationally-relevant-things-considered rationally impermissible,' respectively. Recall that according to MEV, the only 'rationally relevant things' are the available alternatives, the outcomes to which they might lead, the values of these outcomes, and the precise credences given to outcomes conditional on choosing the alternatives that might bring them about. What, then, are the 'rationally relevant things' according to imprecise decision principles? Orthodox decision theory has it that what we should do depends only on our *beliefs* and *desires*. More precisely put, orthodox decision theory endorses something like

> *Normative Folk Psychology (informal).* Which alternatives are permissible or impermissible for decision-makers depends exclusively on their credences and their values.

To state this more precisely, let $\mathsf{CS_R}$ denote the choice-set that contains all and only the alternatives that are *in fact* all-rationally-relevant-things-considered rationally permissible (in contrast to a choice-set $\mathsf{CS_{P1}}$ associated with a particular imprecise decision principle P^1). Let $\mathsf{PS_R}$ denote the prohibited-set that contains all and only the alternatives that are in fact all-relevant-things-considered rationally impermissible (in contrast to a prohibited-set $\mathsf{PS_{P1}}$ associated with a particular imprecise decision principle P^1). So, we can state the claim more formally as

> *Normative Folk Psychology (formal).* Whether any $x \in \mathbf{A} \in \mathfrak{D}$ is such that $x \in \mathsf{CS_R}$ or $x \in \mathsf{PS_R}$ depends only on $\mathbf{R}, \mathbf{V} \in \mathfrak{D}$.

But, how are we supposed to get from $\mathbf{R}, \mathbf{V} \in \mathfrak{D}$ to a verdict about whether $x \in \mathsf{CS_R}$ or $x \in \mathsf{PS_R}$? Well, $\mathbf{R}, \mathbf{V} \in \mathfrak{D}$ suffice to establish the *expected value* of each $x \in \mathbf{A} \in \mathfrak{D}$ relative to every credence function $\mathsf{C_i} \in \mathbf{R}$. So, (as MEV also says) alternatives are declared all-relevant-things-considered rationally permissible or impermissible on the basis of their expected values. Hence, convention in the literature on imprecise decision principles has it that these principles assess alternatives on the basis of their expected values relative to some (not necessarily all) credence functions in the representor.

With all this in hand, let an imprecise decision principle P^1 formally be, for every possible $\mathfrak{D}$, a pair of functions, the first function being from $\mathfrak{D}$ to $\mathsf{CS_{P1}}$, and the second function being from $\mathfrak{D}$ to $\mathsf{PS_{P1}}$, such that for any $x \in \mathbf{A} \in \mathfrak{D}$, whether $x \in \mathsf{CS_{P1}}$ or $x \in \mathsf{PS_{P1}}$ depends on its expected value EV relative to some $C_i \in \mathbf{R} \in \mathfrak{D}$. Next, I'll consider several imprecise decision principles that will help clarify this definition and that will help give the reader a feel for what kind of principles are out there.

§2.5 A Selection of Imprecise Decision Principles

Before considering the imprecise decision principles tentatively adopted by Greaves (2016), Herlitz (2019), and Mogensen (2021)–which are all quite complicated–, consider first the relatively simple imprecise decision principle known as

> *Liberal (informal).* x is permissible iff x maximizes expected value relative to at least one credence function in the representor; otherwise, x is impermissible.[7]

Using subscript LIB to designate Liberal's choice- and prohibited-sets, we can state it formally as

> *Liberal (formal).* For any $\mathfrak{D}$, let $x \in \mathbf{A} \in \mathfrak{D}$ be an element of CS_{LIB} iff x maximizes EV relative to at least one $C_i \in \mathbf{R} \in \mathfrak{D}$; otherwise, x is an element of PS_{LIB}.

To clarify Liberal, let's apply it to the concrete decision-situation presented in the previous subsection. Recall that Nina could choose between making a €100 donation to either the Against Malaria Foundation (x), Animal Charity Evaluators (y), or the Patient Philanthropy Fund (z). The outcomes to which each of these alternatives might lead, their values, and Nina's credences made it so that y maximizes expected value relative to C_1 and x relative to C_2, and z relative to neither, though it has greater expected value than x and y relative to the credence functions according to which these alternatives don't maximize expected value.

Liberal has this to say: Since x and y both maximize expected value relative to (at least) one credence function, both are permissible; since z doesn't maximize expected value relative to any credence function, it's impermissible. Formally, since x maximizes EV relative to $C_1 \in \mathbf{R} \in \mathfrak{D}$, $x \in CS_{LIB}$; since y maximizes EV relative to $C_2 \in \mathbf{R} \in \mathfrak{D}$, $y \in CS_{LIB}$; since there's no $C_i \in \mathbf{R} \in \mathfrak{D}$ relative to which z maximizes EV, $z \in PS_{LIB}$. In sum, $CS_{LIB}=\{x, y\}$ and $PS_{LIB}=\{z\}$.

Next, Greaves (2016, 329; 333–334) tentatively endorsed the imprecise decision principle originally developed and defended by Rinard (2015), who called it

[7] See, e.g., Joyce (2010, 314); Rinard (2015, 5); Greaves (2016, 328); Mogensen (2021, 148).

> *Moderate (informal).* x is permissible iff x maximizes expected value relative to every credence function in the representor; x is impermissible iff x does not maximize expected value relative to any credence function in the representor; otherwise, x is indeterminately permissible.[8]

Using subscript MOD to designate Moderate's choice- and prohibited-sets, we can state this imprecise decision principle formally as

> *Moderate (formal)***.** For any $\mathfrak{D}$, let $x \in \mathbf{A} \in \mathfrak{D}$ be an element of $\mathsf{CS}_{\mathsf{MOD}}$ iff there is no $y \in \mathbf{A} \in \mathfrak{D}$ such that y has greater EV than x relative to some $\mathsf{C}_{\mathsf{i}} \in \mathbf{R} \in \mathfrak{D}$; let x be an element of $\mathsf{PS}_{\mathsf{MOD}}$ iff relative to every C_{i}, there is some y such that y has greater EV than x relative to C_{i}; otherwise, x is neither an element of $\mathsf{CS}_{\mathsf{MOD}}$, nor of $\mathsf{PS}_{\mathsf{MOD}}$.

To clarify Moderate, let's apply it to the running example. Recall again that Nina can choose between making a €100 donation to either the Against Malaria Foundation (x), Animal Chari-

---

[8] Greaves doesn't use the label Moderate, and doesn't directly credit Rinard with developing the imprecise decision principle that she tentatively adopts (though she cites Rinard's paper from which it originates), yet it seems that Moderate is the imprecise decision principle that she has in mind. Here's how Greaves (2016, 329) formulates her favoured view:

"*Supervaluational criterion of permissibility*: It is determinately true that [alternative] A is permissible in [circumstances] C iff there is no other action available in C that has higher expected value with respect to all [credence functions] of the representor. It is determinately false that A is permissible in C iff with respect to each [credence function] of the representor, some other act available in C has higher expected value. Otherwise it is indeterminate whether or not A is permissible in C."

It seems to me that Greaves made a mistake in her formulation (which, of course, I haven't adopted in my statement of Moderate). At the start of the quote above, Greaves writes that "It is determinately true that [alternative] A is permissible in [circumstances] C iff there is no other action available in C that has higher expected value with respect to *all* [credence functions] of the representor" (emphasis added). This entails that: It is determinately true that [alternative] A is permissible in [circumstances] C if A maximizes expected value relative to *at least one* credence function in the representor (cf. Liberal). Meanwhile, Greaves contends that *if* the credence functions in the representor disagree about which alternative(s) maximize(s) expected value, the Supervaluational criterion of permissibility will say that, for any available alternative, it's indeterminate whether it's permissible (see 329; 333). But if the credence functions in the representor disagree about which alternative(s) maximize(s) expected value, then there should be some alternatives that maximize expected value relative to at least one credence function in the representor. So, as Greaves has defined the Supervaluational criterion of permissibility, there should then be some alternatives for which it is determinately true that they are permissible. This is a contradiction. What Greaves presumably had in mind is this: "It is determinately true that [alternative] A is permissible in [circumstances] C iff with respect to all credence functions of the representor, there is no other action available in C that has higher expected value." So defined, if the credence functions in the representor disagree about which alternative(s) maximize(s) expected value, then the Supervaluational criterion of permissibility will indeed say that, for any available alternative, it's indeterminate whether it's permissible. The revised definition of Supervaluational criterion of permissibility is essentially equivalent to Rinard's Moderate.

ty Evaluators (y), or the Patient Philanthropy Fund (z). y maximizes expected value relative to C_1 and x relative to C_2, and z relative to neither, though it has greater expected value than x and y relative to the credence functions according to which these alternatives don't maximize expected value.

Moderate has this to say: Since x and y both maximize expected value relative to some, but not all, credence functions in the representor, both are indeterminately permissible; since z doesn't maximize expected value relative to any credence function, it's impermissible. Formally, since x has greater EV than y relative to C_1, $y \notin CS_{MOD}$; since y has greater EV than x relative to C_2, $x \notin CS_{MOD}$. Since it's not so that relative to every $C_i \in \mathbf{R} \in \mathfrak{D}$, there's some alternative that has greater EV than y, $y \notin PS_{MOD}$. Since it's not so that relative to every $C_i \in \mathbf{R} \in \mathfrak{D}$, there's some alternative that has greater EV than x, $x \notin PS_{MOD}$. Since relative to every $C_i \in \mathbf{R} \in \mathfrak{D}$, there *is* some alternative that has greater EV than z, $z \in PS_{MOD}$. In sum, $CS_{MOD}=\{\}$ and $PS_{MOD}=\{z\}$. Moderate's verdict therefore differs from Liberal's: While these principles agree that z is impermissible, Liberal declares x and y permissible, and Moderate maintains that they're indeterminately permissible.

A clarificatory remark is in order: To say that an alternative is *indeterminately permissible* is to say that it's *indeterminate whether it's true that the alternative is permissible* (a justification of the possibility of indeterminate permissibility in this sense is beyond the scope of this thesis; see Rinard 2015, 1–2). I assume that it's fair to formalize Moderate as saying that indeterminately permissible alternatives are neither elements of its choice-set, nor of its prohibited-set. Greaves (2016) tentatively endorses Moderate precisely because it employs the notion of indeterminate permissibility: This captures the "intuitive sense of cluelessness" that decision-makers feel when confronted with decisions under cluelessness (333).

Consider finally the view to which Mogensen (2021, 146–151) and Herlitz (2019, 13–14) tentatively subscribe,[9] which goes by

[9] More accurately, the view to which Mogensen tentatively subscribes, which happens to be strikingly similar to the view to which Herlitz tentatively subscribes. The differences between Mogensen's and Herlitz's favoured accounts appear to be primarily terminological. The imprecise decision principle endorsed by Herlitz (2019, 13–14), which I'll refer to as Determinate Maximality, can be stated as follows:

Determinate Maximality (informal). x is permissible iff x is *determinately maximal*, i.e., x isn't determinately worse than any y; otherwise, x is impermissible. Furthermore, x is determinately worse than y iff x has lower expected value than y relative to every credence function in the representor; it is indeterminate

Maximality (informal). x is permissible iff there is no y such that y is strictly preferred to x; otherwise, x is impermissible. Furthermore, x is strictly preferred to y iff x has greater expected value than y relative to every credence function in the representor; x is indifferent to y iff x and y have equal expected value relative to every credence function in the representor; otherwise, preferences between x and y are indeterminate.

Using subscript MAX to designate Maximality's choice- and prohibit-sets, we can state this imprecise decision principle formally as

Maximality (formal). For any $\mathfrak{D}$, let $x \in \mathbf{A} \in \mathfrak{D}$ be an element of CS_{MAX} iff there is no $y \in \mathbf{A} \in \mathfrak{D}$ such that y is strictly preferred to x; otherwise, x is an element of PS_{MAX}. Furthermore, x is strictly preferred to y iff has greater EV than y relative to every $C_i \in \mathbf{R} \in \mathfrak{D}$; x is indifferent to y iff x has equal EV to y relative to every $C_i \in \mathbf{R} \in \mathfrak{D}$; otherwise, preferences between x and y are indeterminate.

To clarify Maximality, let's apply it to the running example. Recall that y maximizes expected value relative to C_1 and x relative to C_2, and z relative to neither, though it has greater expected value than x and y relative to the credence functions according to which these alternatives don't maximize expected value.

Maximality has this to say: Under these circumstances, no alternative is strictly preferred to another, and no alternative is indifferent to another. Hence preferences between the alternatives are indeterminate, meaning that Maximality declares x, y, and z all permissible. More formally, for no pair of alternatives there's an alternative that has greater EV than the other

---

whether x is worse than y iff x has greater expected value than y relative to some credence functions in the representor, while y has greater expected value than x relative to other credence functions in the representor.

Using subscript DA to designate Determinate Maximality's choice- and prohibited-sets, we can state this imprecise decision principle formally as

Determinate Maximality (formal). For any $\mathfrak{D}$, let $x \in \mathbf{A} \in \mathfrak{D}$ be an element of CS_{DA} iff there is no $y \in \mathbf{A} \in \mathfrak{D}$ such that x is determinately worse than y; otherwise, x is an element of PS_{DA}. Furthermore, x is determinately worse than y iff y has greater EV than x relative to every $C_i \in \mathbf{R} \in \mathfrak{D}$; it is indeterminate whether x is worse than y iff x has greater EV than y relative to some C_i, and y has greater EV than x relative to other C_i.

Notably, Herlitz supplements Determinate Maximality with a decision principle that states that if Determinate Maximality declares multiple alternatives permissible, then decision-makers can choose between these alternatives on the basis of "their agency and personal preferences, passions and commitments" (16.). Since I don't wish to leave the realm of imprecise decision principles in this thesis, I'll set this aside.

relative to every $C_i \in \mathbf{R} \in \mathfrak{D}$. Hence, no alternative is strictly preferred to another. Furthermore, for no pair of alternatives there's an alternative that has equal EV to the other relative to every $C_i \in \mathbf{R} \in \mathfrak{D}$. Hence, no alternative is indifferent to another. It follows that $CS_{MAX}=\{x, y, z\}$ and $PS_{MAX}=\{\}$. Therefore, Maximality's verdict differs from Liberal's and Moderate's. Recall that according to Liberal, $CS_{LIB}=\{x, y\}$ and $PS_{LIB}=\{z\}$, and that according to Moderate, $CS_{MOD}=\{\}$ and $PS_{MOD}=\{z\}$.

Herlitz (2019) offers little in the way of defence of Maximality. Mogensen (2021, 146–151) basically argues that, while Maximality has its flaws (one of which we'll see in §4.2, and another which we'll see in §5.4), its rivals seem to be even more flawed.

So, then, we're looking for a decision principle that can substitute MEV for the sake of effective altruist rational action guidance (i.e., to help effective altruists identify which actions they should perform, rationally speaking, in order to do the most good) in the face of cluelessness (i.e., uncertainty about outcomes best modelled using imprecise credences). An imprecise decision principle seems a plausible candidate, given that it's meant for rational decision-making when imprecise credences are assigned to outcomes. However, we've now seen that there are several such principles, occasionally giving conflicting prescriptions. Which principle, if any, should effective altruists conform to? The answer to this question isn't clear. If this isn't due to irrationality on our part, it must be possible to be *rationally uncertain* about which imprecise decision principle to adopt.[10] Given that cluelessness seems relevant to decision-making, one might hold that rational uncertainty about imprecise decision principles is, too. This is what *Metanormativism* decrees. If this view is true, and given that effective altruists are indeed rationally uncertain about imprecise decision principles, the appropriate replacement of MEV isn't any such principle, but some other account of choice. To Metanormativism and its possible accounts of choice we turn next.

[10] To be fair, it's not an open and shut case that it can be rational to be uncertain about imprecise decision principles–Wedgwood (2019) has developed an argument to the effect that it cannot be rational to be uncertain about *moral theories* that applies *mutatis mutandis* to imprecise decision principles–but considering that it's highly intuitive, I'll operate on the assumption that it's possible.

§3. Metanormativism about Imprecise Decision Principles

§3.1 Metanormativism about Imprecise Decision Principles, Informally

So, it's possible for decision-makers, such as effective altruists, to be rationally uncertain about which imprecise decision principle to conform to when they face decisions under cluelessness. Problematically, imprecise decision principles sometimes give decision-makers different prescriptions (recall the case with which I've illustrated Liberal, Moderate, and Maximality in §2.5). It follows that it can be rational for effective altruists to be *normatively uncertain.* Someone is normatively uncertain iff she's uncertain about *what she ought to do.* More narrowly, it can be rational for effective altruists to be *decision-theoretically uncertain.* Someone is decision-theoretically uncertain iff she's uncertain about *what she ought rationally to do.* This makes decision-theoretic uncertainty a subtype of normative uncertainty. It's noteworthy that normative uncertainty has other subtypes: There's *moral uncertainty* (someone is morally uncertain iff she's uncertain about what she should morally do) and *prudential uncertainty* (someone is prudentially uncertain iff she's uncertain about she should do for her own sake).

Various authors have recently contended that normative uncertainty in general, or decision-theoretic, moral, or prudential uncertainty in particular, should be accounted for in decision-making, a position sometimes labelled Metanormativism, Uncertaintism, or Normative Internalism (see, e.g., Nozick 1995, 34–50; MacAskill 2014; 2016b; Tarsney 2019a; Sepielli 2014; Greaves and Ord 2017; MacAskill and Ord 2020; MacAskill, Bykvist, and Ord 2020; Podgorski 2020; Trammell 2021; Riedener 2021; MacAskill *et al.* 2021; Mogensen 2021, fn. 16). The bulk of the literature has centred on moral uncertainty, but the general idea is that there's some account of choice that takes one or more of these forms of uncertainty into consideration in giving us prescriptions. The most simple account of this sort–known as My Favourite Theory (MFT)–simply says that, if confronted with normative (decision-theoretic, moral, prudential) uncertainty, we should act in accordance with the normative (decision-theoretic, moral, prudential) view that we find most plausible.

While researchers sympathetic to the view that one or more of these uncertainties is relevant to decision-making are typically more interested in *teasing out the implications* of this position than in *defending* it, various defences have been presented. To give just a few examples,[11] it has been pointed out that this position has *intuitive appeal.* Furthermore, it may seem *arbi-*

[11] Adapted from MacAskill and Ord (2020, 329–332; 339).

trary to take uncertainty about outcomes into consideration when making decisions–as imprecise decision principles do–, but to ignore uncertainty about normative (decision-theoretic, moral, prudential) views. Thirdly, normative (decision-theoretic, moral, prudential) views about which we're uncertain don't seem sufficiently *action guiding*, for how can a view of which we don't know that we ought to follow it help us identify which actions to perform? Considering the value that effective altruism gives to action guidance (see §2.1), it seems that Metanormativism and effective altruism fit nicely.

Some philosophers, however, remain unmoved by these sorts of arguments, and deny that normative uncertainty in general, or decision-theoretic, moral, or prudential uncertainty in particular, should be accounted for in decision-making, a view sometimes labelled Actualism or Normative Externalism (see Weatherson 2014; 2019; Harman 2015; Hedden 2016).

In the present context, the position of interest is a restricted form of Metanormativism, namely

> *Metanormativism about Imprecise Decision Principles.* Rational uncertainty about imprecise decision principles should rationally be accounted for in decision-making.

If Metanormativism about Imprecise Decision Principles (abbreviated as Metanormativism) is true, then there is some account of rational decision-making that takes (among other things) decision-makers's rational uncertainty about imprecise decision principles into consideration when giving them prescriptions. It seems to me that the Metanormativist account of rational choice can take one of two important, general forms.

First, it might take the form of a Unique Account, i.e., a unique second-order decision principle, sometimes also called a meta decision theory, that decision-makers should under all circumstances rely on to factor uncertainty about imprecise decision principles into their decision-making. A position such as this has been defended by Tarsney (2019). Roughly, a second-order decision principle is an account that specifies which alternatives are rationally permissible or impermissible on the basis of (i) precise or imprecise credences given to imprecise decision principles and (ii) the permissibility or impermissibility of the available alternatives according to imprecise decision principles to which credence is given. MFT is an example of a second-order decision principle.

Second, the Metanormativist account might take the form of a Hierarchy Account, i.e., an infinite hierarchy of second- and higher order decision principles to which decision-makers give credence. To see what this means, note that in case decision-makers are uncertain about second-order decision principles, the Metanormativist might admit that a single second-order decision principle cannot do the trick (i.e., that there's no adequate Unique Account). She might claim that, instead, the decision-maker should resort to a third-order decision principle to come to a decision. Roughly, a third-order decision principle is an account that specifies which alternatives are rationally permissible or impermissible on the basis of (i) precise or imprecise credences given to second-order decision principles and (ii) the permissibility of impermissibility of the available alternatives according to the second-order decision principles to which credence is given. Uncertainty about third-order decision principles should in turn be handled by means of a fourth-order decision principle, and so on. Nonetheless, the Metanormativist might hold that, in the face of such infinite regress, rational decision-making isn't (necessarily) impaired. Positions of this sort have been taken up by Sepielli (2014), MacAskill (2014, 217–219), and Trammell (2021).

With regards to effective altruism, it's fair to presume that much is at stake with the decision about whether to endorse Metanormativism, be it in the form of a Unique Account or a Hierarchy Account. Adding information about decision-makers's uncertainty about imprecise decision principles to the decision process can seemingly have a considerable impact on which alternatives are declared permissible or impermissible. Various prominent and influential effective altruists, most notably MacAskill, Tarsney, Ord, Greaves, and Mogensen, have publicly endorsed Metanormativism, or positions affiliated with it. (Given that the need for action guidance is one of the motivations for endorsing Metanormativism in its unrestricted form, this is perhaps unsurprising). Due to such endorsements, Metanormativism pertaining to moral theories, i.e., the view that moral uncertainty should be accounted for in decision-making, has already taken hold in effective altruism. To give just one example of the influence of this view: Effective altruism's aim of 'doing the most good' is often qualified to account for moral uncertainty about utilitarianism or maximizing consequentialism. To this end, it's added that alternatives that violate rights or deontological constraints should generally not be chosen, even if choosing them does the most good (see, e.g., Sebo and Paul 2019, 54; MacAskill and Pummer 2020, 5; Berkey 2021, 98-99). So, Metanormativism (i.e., Metanormativism about imprecise decision principles) could go on to exert influence on effective altruist decision-making as well (if it hasn't already)–unduly, or so I'll argue. By arguing against both general types of Metanormativist account, I'll argue that there's no plausible Metanormativist account of rational decision-making, from which it follows that Metanormativism is false.

For this purpose, I'll first extend the formal framework presented in §2.4 so that we can talk about Unique and Hierarchy Accounts more precisely. I'll do this in the next subsection. But let me note here that while Unique and Hierarchy Accounts are, in my view, the most important types of Metanormativist accounts of rational choice, they aren't the only conceivable ones: It's possible to make adjustments to Unique and Hierarchy Accounts as I've described them, resulting in alternative Metanormativist accounts of choice. Nevertheless, for my argument against Metanormativism, it suffices that I only consider Unique Accounts and Hierarchy Accounts, as I've described them. Let me explain.

As I've defined the notion of second-order decision principle, it specifies which alternatives are rationally permissible or impermissible on the basis of (i) precise or imprecise credences given to imprecise decision principles and (ii) the permissibility or impermissibility of the available alternatives according to imprecise decision principles to which credence is given. It's possible to rethink second-order decision principles, like so: A second-order decision principles specifies which alternatives are rationally permissible or impermissible on the basis of (i) precise or imprecise credences given to imprecise decision principles, (ii) *precise or imprecise credences given to second-order decision principles*, and (ii) the permissibility or impermissibility of the available alternatives according to imprecise decision principles *and second-order decision principles* to which credence is given. Similarly, third- and higher order decision principles needn't only take into account the order below, but could be formulated so that uncertainty about these principles themselves feeds back into them.

Moreover, we could allow *imprecise decision principles* to do something similar. As I've defined imprecise decision principles, they factor in credences ranging over outcomes o_1, o_2, …, o_n. But what if o_1 is an outcome in which Maximality is true and o_2 an outcome in which Liberal is true? If we'd allow for this, an imprecise decision principle would be a Metanormativist account of sorts.

With regards to Hierarchy Accounts, what about decision principles that aren't part of the hierarchy, but that are somehow external to it, ranging over all principles jointly build the hierarchy? We could think of Hierarchy Accounts as potentially involving such decision principles.

As for Unique Accounts, why not go *pluralist*? That is, why not claim that in such and such second-order decision-situations, decision-makers should always use some unique second-order decision principle, while in such and such other second-order decision-situations, decision-makers should always use some other second-order decision principle? Or (alternatively) why not claim that there isn't a unique second-order decision principle, but rather, say, a unique *third-* or *ninth-*order decision principle? That is, why not claim that there is some third- or ninth-order decision principle that should always be used to account for uncertainty about decision principles at the order below? The resulting account of choice would resemble a kind of finite Hierarchy Account.

Finally (although, to be sure, I'm not aiming to be exhaustive), in the literature on decision-making under normative uncertainty in general, or decision-theoretic, moral, or prudential uncertainty in particular, it's conventional for higher order decision principles to give decision-makers prescriptions on the basis of claims about the *choice-worthiness* of the available alternatives according to lower order decision principles (see, e.g., Ross 2006; MacAskill 2016a; MacAskill and Ord 2020). So, it would be more in line with the literature to define second-order decision principles as specifying which alternatives are rationally permissible or impermissible on the basis of (i) precise or imprecise credences given to imprecise decision principles and (ii) the *choice-worthiness* of the available alternatives according to imprecise decision principles to which credence is given. The choice-worthiness of an alternative x according to an imprecise decision principle P^1 is, roughly, just how badly P^1 wants the decision-maker to choose x.

So, then, there are various ways in which we can extend and adjust Unique and Hierarchy Accounts. As far as I can tell, introducing such adjustments will not save Metanormativism from the objections to come. That is, with suitable changes, the arguments that I'll develop will also apply to adjusted–and perhaps superior, though certainly more complicated–Metanormativist accounts. It's, however, beyond the scope of this thesis to show this. So, for reasons of space, I'll set alternative Metanormativist accounts aside (although the notion of 'choice-worthiness' will make a reappearance in §§4.2-4.4, I'll briefly argue against Unique Accounts that make use of a unique third- or higher order decision principle instead of a unique second-order decision principle in §5.2, and I'll briefly allow there to be decision principles external to the infinite hierarchy, ranging over all principles that are part of the hierarchy, in §5.3). Next, I'll extend the formal framework presented in §2.4 so that we can characterize Unique and Hierarchy Accounts in a precise manner.

§3.2 A Formalism for Metanormativism about Imprecise Decision Principles

Similarly to how I've defined imprecise decision principles as pairs of functions from decision-situations (see §2.4), I'll define *second-order* decision principles as pairs of functions from *second-order* decision-situations; *third-order* decision principles as pairs of functions from *third-order* decision-situations; and so on.

Informally, for any $n>1$, let an n*th-order decision-situation* be any case in which a decision-maker has to come to a decision under cluelessness and rational uncertainty about $(n\text{-}1)^{\text{th}}$-order decision principles, $(n\text{-}2)^{\text{th}}$-order decision principles, $(n\text{-}3)^{\text{th}}$-order decision principles, and so on, such that the lowest order is the 1^{st}. So, informally, a *second-order decision-situation* is a case in which a decision-maker has to come to a decision under cluelessness and rational uncertainty about first-order decision principles, i.e., imprecise decision principles; a *third-order decision-situation* is a case in which a decision-maker has to come to a decision under cluelessness, rational uncertainty about second-order decision principles, and rational uncertainty about imprecise decision principles; and so on.

Formally, for any $n>1$, let an n^{th}-order decision-situation $\mathfrak{D}^n$ be a tuple $(\mathbf{D}, \mathbf{A}, \mathbf{O}, \mathbf{R}^n, \mathbf{V}, \mathbf{P}^{(n\text{-}1)})$, where $\mathbf{D}$, $\mathbf{A}$, $\mathbf{O}$, and $\mathbf{V}$ are as before (see §2.4, or, for your convenience, this footnote);[12] $\mathbf{P}^{(n\text{-}1)}$ is a set of mutually exclusive and jointly exhaustive $(n\text{-}1)^{\text{th}}$-order decision principles $P^{(n\text{-}1)}$, such that $|\mathbf{P}^{(n\text{-}1)}|\geq 2$; $\mathbf{R}^n$ is subset of $\mathbf{D}$'s representor, such that $\mathbf{R}^n$ contains only the credences that n^{th}-order decision principles might deem relevant for coming to a decision; that is, $\mathbf{R}^n$ is such that each credence function $C_i \in \mathbf{R}^n$ assigns only rationally admissible numerical values to, for each $C_i \in \mathbf{R}^2$ and for each $(n\text{-}1)^{\text{th}}$-order decision principle $P^{(n\text{-}1)} \in \mathbf{P}^{(n\text{-}1)}$, $C_i(P^{(n\text{-}1)})$.

[12] $\mathbf{D}$ is a decision-maker; $\mathbf{A}$ is a set of available alternatives, i.e., a subset of the set of all possible alternatives $\mathfrak{A}$, such that $|\mathbf{A}|\geq 2$; $\mathbf{O}$ is a subset of the set of all possible outcomes $\mathfrak{O}$, such that it contains all and only the outcomes that might be brought about by the alternatives in $\mathbf{A}$; so, for each alternative $x \in \mathbf{A}$, a subset of $\mathbf{O}$ contains the mutually exclusive and jointly exhaustive outcomes that x might bring about; $\mathbf{V}$ is a value function, such that for each $O_i \in \mathbf{O}$, $\mathbf{V}(O_i)$ is assigned a rationally admissible numerical value, where the greater the numerical value assigned to $\mathbf{V}(O_i)$, the greater the value (i.e., moral value) of O_i, such that a value function $\mathbf{V}$ can be transformed into a value function $\mathbf{V}'$ by means of a positive linear transformation.

To illustrate, suppose that Nina (from §§2.4-2.5) is aware of her predicament. That is, she's aware that Liberal, Moderate and Maximality disagree about the permissibility and impermissibility of the alternatives available to her. Recall that according to Liberal, $CS_{LIB}=\{x, y\}$ and $PS_{LIB}=\{z\}$, according to Moderate, $CS_{MOD}=\{\}$ and $PS_{MOD}=\{z\}$, and according to Maximality, $CS_{MAX}=\{x, y, z\}$ and $PS_{MAX}=\{\}$. Let's say, then, that Nina is in a second-order decision-situation $\mathfrak{D}^2$. She's the decision-maker $\mathbf{D}$, the available alternatives remain making a €100 donation to either the Against Malaria Foundation (x), Animal Charity Evaluators (y), or the Patient Philanthropy Fund (z) (in which case $\mathbf{A}=\{x, y, z\}$ and $|\mathbf{A}|=3$). The outcomes to which these alternatives might lead and their values are as before (i.e., $\mathbf{O}=\{O_1, O_2, O_3, O_4, O_5, O_6\}$, with $\mathbf{V}$ such that each $O_i \in \mathbf{O}$ is assigned a value). Suppose that the only 'lower order' decision principles to which Nina gives credence–all of which must by definition be first-order, or imprecise, decision principles–are Liberal, Moderate, and Maximality. Hence we can say that $\mathbf{P}^{(n\text{-}1)}=\{$Liberal, Moderate, Maximality$\}$. Finally, suppose that her credences with respect to these principles can be represented by means of two credence functions, C_3 and C_4 (due to which $\mathbf{R}^2=\{C_3, C_4\}$). Credence function C_3 is such that C_3(Liberal)$=0.5$; C_3(Moderate)$=0.3$; C_3(Maximality)$=0.3$. Credence function C_4 is such that C_4(Liberal)$=0.2$; C_4(Moderate)$=0.5$; C_4(Maximality)$=0.3$.

Informally, for any $n>1$, let an nth*-order decision principle* P^n be an account that specifies, for every possible n^{th}-order decision-situation $\mathfrak{D}^n$, which alternatives are permissible or impermissible for the decision-maker on the basis of (i) precise or imprecise credences given to $(n\text{-}1)^{\text{th}}$-order decision principles $P^{(n\text{-}1)}$, and (ii) the permissibility or impermissibility of the available alternatives according to these $(n\text{-}1)^{\text{th}}$-order decision principles $P^{(n\text{-}1)}$.

Before formally defining the notion of n^{th}-order decision principle P^n, two remarks are in order. First, just as with imprecise decision principles P^1, the permissible alternatives according to an n^{th}-order decision principle P^n go into its choice-set CS_{P^n}, i.e., a subset of $\mathbf{A} \in \mathfrak{D}^n$ associated with P^n. The impermissible alternatives go into its prohibited-set PS_{P^n}, also a subset of $\mathbf{A} \in \mathfrak{D}^n$ that is associated with P^n. As before, alternatives can't be both permissible and impermissible, meaning that if $x \in PS_{P^n}$, then $x \notin CS_{P^n}$, and if $x \in CS_{P^n}$, then $x \notin PS_{P^n}$.

Second, note that we need to forge some sort of *connection* between decision-situations at different orders. Consider Nina again. Given her cluelessness in her (first-order) decision-situation $\mathfrak{D}$, she wants to come to a decision using an imprecise decision principle. Regrettably, as we have just seen, she's uncertain about them. Hence she has 'moved up' one order, to a second-order decision-situation $\mathfrak{D}^2$, meaning that she's in the position to use a second-order decision principle to deal with her uncertainty. As I've informally defined the concept of 'second-order decision principle,' it will specify which alternatives are permissible or impermissible for Nina *in this second-order decision-situation* $\mathfrak{D}^2$ on the basis of claims about the permissibility or impermissibility of alternatives according to the imprecise decision principles *in her first-order decision-situation* $\mathfrak{D}$. So, to make sense of the notion of 'second-order decision principle,' we'll need to forge some connection between the initial (first-order) decision-situation $\mathfrak{D}$ and the second-order decision-situation $\mathfrak{D}^2$. The same goes for higher order decision-situations: Each decision-situation should be connected to the one that 'came before,' i.e., for any n>2, every n^{th}-order decision-situation $\mathfrak{D}^n$ should be connected to an $(n\text{-}1)^{th}$-order decision-situation $\mathfrak{D}^{(n-1)}$. I'll not attempt to spell out necessary and sufficient conditions for decision-situations being connected to one another, but take it that the notion of 'connection' is sufficiently clear for present purposes.

Now, formally, for any n>1, let an n^{th}-order decision principle P^n be, for every possible $\mathfrak{D}^n$, a pair of functions, the first being from $\mathfrak{D}^n$ to CS_{Pn}, and the second being from $\mathfrak{D}^n$ to PS_{Pn}, such that for any $x \in \mathbf{A} \in \mathfrak{D}^n$, whether $x \in CS_{Pn}$ or $x \in PS_{Pn}$ depends only on (i) for some $P^{(n-1)} \in \mathbf{P}^{(n-1)} \in \mathfrak{D}^n$ and for some $C_i \in \mathbf{R}^n \in \mathfrak{D}^n$, $C_i(P^{(n-1)})$, and (ii) for some $P^{(n-1)}$, $CS_{P(n-1)}$ or $PS_{P(n-1)}$ such that these choice- or prohibited sets are associated with the $(n\text{-}1)^{th}$-order decision-situation $\mathfrak{D}^{(n-1)}$ to which $\mathfrak{D}^n$ is connected. (We'll see proper, formally spelled out examples of n^{th}-order decision principles in §4.3, but since these are somewhat elaborate, they are best reserved for later).[13]

[13] The very astute reader might have noticed that some of the assumptions that I've made so far entail a paradox. Recall from §2.3 that the terms 'permissible' and 'impermissible' should be understood as 'all-rationally-relevant-things-considered rationally permissible' and 'all-rationally-relevant-things-considered rationally impermissible,' respectively. So, n^{th} -order decision principles, for any n>1, specify whether alternatives are all-rationally-relevant-things-considered rationally permissible or impermissible. Imprecise decision principles specify whether alternatives all-rationally-relevant-things-considered rationally permissible or impermissible, too. Any decision-maker who finds herself in a third- or higher order decision-situation will therefore give credence to (i) a set of mutually exclusive and jointly exhaustive imprecise decision principles that specify what she ought to all-rationally-relevant-things-considered do and (ii) a set of mutually exclusive and jointly exhaustive second-order decision principles that specify what she ought to all-rationally-relevant-things-considered do. This entails that she should be certain that (a) one and only one imprecise decision principle correctly specifies what she should all-rationally-relevant-things-considered do and that (b) one and only one second-order decision principle also correctly specifies what

We now have the tools to state second-order decision principles precisely, allowing us to evaluate Unique Accounts, i.e., single second-order decision principles that decision-makers should purportedly always conform to for the sake of dealing with uncertainty about imprecise decision principles in their decision-making. Furthermore, we now have the tools to evaluate Hierarchy Accounts, i.e., infinite hierarchies of second- and higher order decision principles to which decision-makers give credence. In the next section, I'll argue that there's no plausible Unique Account. In the section after the next, I'll argue that there's no plausible Hierarchy Account.

---

she should all-rationally-relevant-things-considered do. That's impossible–so, something has got to give. The reader will likely anticipate my response: This paradox casts doubt on the whole project of decision-making on the basis of rational uncertainty about imprecise decision principles, i.e., it casts doubt on Metanormativism. We should do away with the concepts of 'second-order decision principle,' 'third-order decision principle,' and so on. The Metanormativist will presumably remain unpersuaded by this objection as it's now stated. MacAskill *et al.* (2021, 328)–who endorse a position similar to Metanormativism–are aware of the sort of tension that I've pointed to. The authors do not attempt to resolve it, but proceed on the assumption that a solution is out there, one that preserves Metanormativism. It's beyond the scope of the paper to survey the possible Metanormativist replies to the paradox. The argument that I'll develop in the remainder of this thesis makes, I think, for a more powerful objection to Metanormativism, so I'll concentrate on that. Thus, in the remainder of this thesis, I'll grant the Metanormativist that the contradiction can be resolved without giving up op Metanormativism. I've done my best to frame the upcoming argument against Metanormativism in such a way that it's compatible with–relatively–plausible Metanormativist responses to the paradox.

§4. Against Unique Accounts

§4.1 Set Up

In this section, I'll argue that there's no adequate Unique Account, i.e., that there's no single second-order decision principle that decision-makers should exclusively rely on to take their uncertainty about imprecise decision principles into account in their decision-making. As far as I'm aware, among adherents of (positions similar to) Metanormativism, it's not a popular view that there indeed is such a second-order decision principle. Various authors concerned with decision-making under normative uncertainty in general, or moral, prudential, or decision-theoretic uncertainty in particular, have defended particular second-order decision principles (see, e.g., Gracely 1996; Gustafsson and Torpman 2014; MacAskill 2016a; MacAskill and Ord 2020). But, what seems to be at stake in these debates is which second-order decision principle is *true*, and claiming that some second-order decision principle is true isn't equivalent to claiming that it should always be relied upon. The Metanormativist might hold that there's *also* a true third-order decision principle, which should be used in cases of uncertainty about second-order decision principles. That said, the position that there's an adequate Unique Account has at least one defendant, namely Tarsney (2019). Furthermore, as indicated in the previous section, it's the only serious alternative to Hierarchy Accounts, so if the Metanormativist doesn't wish to endorse one of those (perhaps due to my objection from the next section), she's forced to accept that there's an adequate Unique Account. So, to argue against Metanormativism, it's worth arguing against Unique Accounts.

Here's how I'll proceed. To argue that there's no single second-order decision principle to which decision-makers should always resort, it'll be helpful to get our hands on some concrete second-order decision principles (this will also prove helpful in arguing against Hierarchy Accounts in the next section, so it serves a dual purpose). As I've just noted, various authors have defended particular second-order decision principles. For the most part, these are second-order decision principles for decision-making under *moral uncertainty* (i.e., uncertainty about what should be morally done), but they serve our purpose well enough. So, in §4.2, I'll briefly survey the literature on second-order decision principles for the sake of decision-making under moral uncertainty.

Notably, discussion about decision-making under normative uncertainty in general, or moral, prudential, or decision-theoretic uncertainty in particular, has proceeded on the assumption that decision-makers assign *precise* rather than *imprecise* credences. To my knowledge, the possibility of imprecise credences has so far only been acknowledged in passing (see, e.g., MacAskill and Ord 2020, 329; MacAskill 2016a, fn. 52; see also Mogensen 2021, fn. 16). This means that it has generally been assumed that decision-makers assign precise rather than imprecise credences to normative (moral, decision-theoretic, prudential) views. Consequently, the second-order decision principles that have been developed and defended typically cannot handle cases in which decision-makers assign imprecise credences to normative (moral, decision-theoretic, prudential) views. It's implausible to suppose that decision-makers can only assign precise credences to imprecise decision principles (these principles being relevant *only* if imprecise credences are or should be assigned). So, if there's to be a unique second-order decision principle for the sake of taking uncertainty about imprecise decision principles into account in decision-making, it should be able to accommodate imprecise credences.

In §4.3, I'll develop such second-order decision principles. Each second-order decision principle that I'll propose is essentially a hybrid of (i) a second-order decision principle from the literature on moral uncertainty (among other sources) and (ii) an *imprecise decision principle* of the sort discussed in §§2.4-2.5. Given that it's possible for decision-makers to be rationally uncertain about imprecise decision principles, it should therefore be possible for decision-makers to be rationally uncertain about these second-order decision principles.

As I'll point out in §4.4, this means that, while Metanormativism claims that rational uncertainty about imprecise decision principles should be accounted for in decision-making, Unique Accounts maintain that rational uncertainty about second-order decision principles inspired *by these very same imprecise decision principles shouldn't be accounted for in decision-making*. This is an unpalatable implication. To round off, I'll discuss Tarsney's (2019) Unique Account. By then, we'll have all the tools we need to show that, at least in the present context, it will not do.

§4.2 Decision-Making under Moral Uncertainty

In this subsection, I'll briefly survey the literature on decision-making under moral uncertainty, with an eye on proposals for second-order decision principles. It'll be useful to centre on the discussion about MFT, which can, in this context, be defined as

My Favourite Theory. An alternative *x* is permissible iff *x* is morally permissible according to the (a) moral view that the decision-maker finds most plausible.

MFT is a commonsensical approach, as applied to dealing with moral uncertainty, uncertainty about imprecise decision principles, or any form of uncertainty, really. Many people presumably unconsciously adhere to something like MFT. However, I'm not singling out MFT because of its plausibility–as we'll see, it's anything but plausible. MFT faces substantial criticism in the context of decision-making under moral uncertainty (most objections are detailed and discussed in Gustafsson and Torpman 2014; see also Gustafsson forthcoming). Hence, it's an incredibly unpopular theory for decision-making under moral uncertainty among the experts in the field. I'll show for three objections to MFT from the literature on moral uncertainty that each can be used to support an alternative second-order decision principle to take moral uncertainty into account in one's decision-making. So, we'll end up with a small collection of second-order decision principles.

There's another reason that I'm concentrating on MFT. It's not essential for the argument against Unique Accounts, but nevertheless interesting in the present context. Recall that Greaves (2016), Herlitz (2019), and Mogensen (2021) share our interest in effective altruist decision-making under cluelessness. Mogensen in particular delves quite deep into the practical implications of the presence of cluelessness on effective altruist decision-making. In his discussion, Mogensen implicitly relies on MFT (though as applied to imprecise decision principles rather than moral views). Given his implicit reliance on MFT, we'll be evaluating Mogensen's investigation into effective altruist decision-making under cluelessness by evaluating this second-order decision principle.

To see that Mogensen implicitly relies on MFT, note that he invokes Maximality to assess effective altruist decision-making. At least as far as this imprecise decision principle is concerned, Mogensen concludes after a lengthy argument that, roughly put, for many decisions about what real-life charitable organizations to donate to, effective altruists are permitted to support *any* organization (contrary to effective altruist orthodoxy) (151–157). That said, Mogensen acknowledges that it's an open question whether Maximality is correct (146–151). For instance, as Mogensen (2021, 155–156) points out, Maximality fails to meet

Sen's Condition β. If x and y are both permissible whenever they are the only available alternatives, then in the presence of a third alternative z, x is permissible iff y is permissible.[14]

Nonetheless, Mogensen claims that "we cannot rule out [Maximality]. As a result, we ought to avoid drawing any conclusions that are inconsistent with it" (Mogensen 2021, 151) and "we cannot rule out the rational permissibility of acts that are evaluated as such by the theory" (fn. 15). If so, we cannot rule out the 'permissive' conclusion that, roughly, for many decisions about what charitable organizations to donate to, effective altruists are permitted to donate to any organization, and we should avoid concluding otherwise.

Mogensen acknowledges the relevance of decision-making under uncertainty about imprecise decision principles for his argument, but sets it aside for reasons of space (fn. 16). Despite this, it's best to understand his line of reasoning as accounting for such uncertainty and, more specifically, as involving appeal to MFT. This is because, when taken at face value, it's peculiar to claim that "we cannot rule out [Maximality]. As a result, we ought to avoid drawing any conclusions that are inconsistent with it." Besides Maximality, there are many more imprecise decision principles that cannot be ruled out (e.g., Moderate, Liberal). These principles will occasionally make recommendations that are inconsistent with Maximality, and this simply cannot be avoided (see, e.g., the case that I used to illustrate Liberal, Moderate, and Maximality in §2.5). However, perhaps it's not the mere fact that "we cannot rule out" Maximality that makes it so that "we ought to avoid drawing any conclusions that are inconsistent with it." Mogensen gives the impression that he finds Maximality the most attractive imprecise decision principle, saying that we may want to "prefer [it] on balance" (146). Perhaps the claim that "we ought to avoid drawing any conclusions that are inconsistent with it" rests on Maximality being the most plausible imprecise decision principle in light of the available evidence. If this interpretation is right, then Mogensen's argument for the 'permissive' conclusion that, roughly, for many decisions about what charitable organizations to donate to, effective altruists are permitted to support any organization, is an instance of reasoning by means of MFT.

So, let's turn to the objections to MFT.

[14] My formulation of the condition is somewhat simplified. See also Sen (2017, 63–64).

§4.2.1 Objection 1 to MFT: My Favourite Alternative

The first objection to MFT runs roughly as follows (see Gustafsson and Torpman 2014, 165 and the references therein). Suppose that a decision-maker gives precise credences to *ninety-one* different moral views. Which exact moral views these are (e.g., utilitarianism, Kantianism, care ethics) is irrelevant for present purposes. What matters is that she gives credence 0.1 to a moral view that tells her that only alternative *x* is permitted. In addition, she gives credence 0.01 to each of the ninety alternative moral views, all of which tell her to choose *y*. Hence, *y* is by far the decision-maker's *Favourite Alternative*: She should give credence 0.1 to the proposition that *x* is permissible, and the proposition that *y* is permissible should receive credence 0.99. Even so, MFT tells her to choose *x*, since she gives greater credence to the moral view that deems *x* permissible than to each of the individual moral views that deem *y* permissible. This is intuitively implausible. Hence, MFT is implausible.

As Gustafsson and Torpman (2014, 165) realized, this objection presupposes something like

> *My Favourite Alternative.* *x* is permissible iff *x* is the (an) alternative that is most likely permissible.[15]

This is an alternative second-order decision principle.

§4.2.2 Objection 2 to MFT: Violation of a Dominance Principle

The second objection runs as follows (see Gustafsson and Torpman 2014, 169). Consider a case in which the moral view that is given highest credence permits alternatives *x* and *y*. Meanwhile, all other moral views to which the decision-maker gives credence permit only *y*. This makes *x* a so-called *dominated* alternative. Even though *x* might indeed be permissible, the decision-maker is certain that *y* is. So, she should choose *y*. However, MFT doesn't require her to do so; she's permitted to choose *x* as well, since *x* is permitted by the moral view to which she gives most credence. Hence MFT violates the following plausible

> *Dominance Principle.* *x* is impermissible if *x* is dominated, i.e., *x* is impermissible according to some moral views to which the decision-maker gives credence and there

[15] The view under consideration is known as *My Favourite Option*, not My Favourite Alternative. Since I use the term 'alternative,' not 'option,' the change is suitable.

exists some *y* that is permissible according to all moral views to which the decision-maker gives credence.

To rescue MFT from this objection, Gustafsson and Torpman (170) suggest a variation on it, which can (setting aside irrelevant complications) be stated as follows

Dominance-My Favourite Theory. *x* is permissible iff *x* is morally permissible according to the (a) moral view that the decision-maker finds most plausible, such that *x* is not dominated.

So we have a third second-order decision principle.

§4.2.3 Objection 3 to MFT: Intertheoretic Comparisons of Choice-Worthiness

In a nutshell, the final objection that I'll consider is that it's (sometimes) possible to make *intertheoretic comparisons of choice-worthiness* across moral views, that these are relevant in coming to a decision, but aren't exploited by MFT (see, e.g., Gustafsson and Torpman 2014, 160–65 and the references therein).

I've already mentioned the notion of choice-worthiness in §3.1 in relation to imprecise decision principles, defining the choice-worthiness of an alternative *x* according to an imprecise decision principle P^1 as a measure of how badly P^1 wants the decision-maker to choose *x*. A similar definition might apply to moral views. In fact, we might even say that choice-worthiness is comparable across moral views. I'll explain. Certain deontological theories condemn murder outright. According to some such theories, killing two people isn't worse than killing one–it's simply forbidden to kill anyone. So, we can say that according to such moral views, these two alternatives are *equally choice-worthy*. Utilitarianism disagrees: Depending on how we spell out the case it could say that killing two people is twice as bad as killing one, giving the latter *greater choice-worthiness*. If intertheoretic comparisons of choice-worthiness across these moral theories are possible, we can say that the distance in choice-worthiness with regards to killing one person and killing two is greater according to utilitarianism than it is according to the deontological theory. There's a lively debate on whether we can indeed make intertheoretic comparisons of choice-worthiness across moral theories (Tarsney 2018 has suggested a promising approach that seemingly works for some, though not all, moral theories).

Let's suppose that we can indeed make intertheoretic comparisons of choice-worthiness, across at least *some* moral views. If so, then this type of information seems relevant for MFT to take into account. It doesn't, though, and this counts against MFT.

There's, however, a second-order decision principle that does take intertheoretic comparisons of choice-worthiness across moral views into account. It's basically a variation on MEV, adapted so as to be applicable to decisions under moral uncertainty. It's by far the most popular second-order decision-making in the domain of decision-making under moral uncertainty (see, e.g., Oddie 1994; Ross 2006; Sepielli 2009; MacAskill and Ord 2020; see also Wedgwood 2013). It goes by various names, most often

> *Maximize Expected Choice-Worthiness (informal).* x is permissible iff x maximizes expected choice-worthiness.

The *expected choice-worthiness* of x is defined similarly to its expected value, with the most substantive change being that the value function $\mathbf{V}$ (see §2.3) is replaced with a *choice-worthiness function* CW. A choice-worthiness function CW associated with a moral view captures how badly this view wants alternatives to be chosen, quantitatively. Formally, a choice-worthiness function CW is a function from a set of alternatives to $\mathbb{R}$, where the greater the numerical value assigned to $\mathrm{CW}(x)$, the greater the choice-worthiness of x, such that choice-worthiness is measured on an interval scale, i.e., a choice-worthiness function CW can be transformed into a choice-worthiness function CW′ by means of a positive linear transformation (cf. MacAskill 2016a, 971). Assuming that it's possible to make intertheoretic comparisons of choice-worthiness across moral theories, it's possible to normalize their choice-worthiness functions, i.e., to place them on a *common scale*. The expected choice-worthiness EC of x, then, is

$$\mathrm{EC}(x) = \sum_{i=1}^{n} \mathrm{C(M)CW}_M(x)$$

where n is the number of moral views to which the decision-maker gives credence, C(M) is the credence the decision-maker gives to moral view M, and $\mathrm{CW_M}(x)$ is the choice-worthiness of x according to M. Precisely put, then, Maximize Expected Choice-Worthiness (MEC) says

> *Maximize Expected Choice-Worthiness (formal).* x is permissible iff there's no available alternative y such that $\mathrm{EC}(y)>\mathrm{EC}(x)$.

So, we have a fourth second-order decision principle.

§4.2.4 Conclusions about MFT

These, then, are three objections to MFT from the literature on decision-making under moral uncertainty. Collectively, these objections show that MFT is wholly inadequate. Claiming that MFT is certainly false would be overstating the case, but the objections give decision-makers good grounds to grant MFT little credence. The failure of MFT undermines Mogensen's (2021) 'permissive' conclusion pertaining to effective altruist decision-making in the face of cluelessness, i.e., that, roughly, for many decisions about what real-life charitable organizations to donate to, effective altruists are permitted to support any organization. More to the point, via the objections we have obtained a total of four second-order decision principles.

§4.3 A Selection of Second-Order Decision Principles

In this subsection, I'll build on the second-order decision principles for decision-making under moral uncertainty laid out in §4.2 to develop second-order decision principles that can be used to take uncertainty about imprecise decision principles into account. Any remotely plausible Unique Account should be able to handle uncertainty about imprecise decision principles that is best modelled using precise or imprecise credences. After all, a Unique Account consists of a single second-order decision principle that should be applied to *all cases* in which decision-makers have to come to decisions in the face of uncertainty about imprecise decision principles. Supposing that all such cases involve either only precise or only imprecise credences given to imprecise decision principles is implausible.

Let's start with MFT, which can in the present context be defined as

> *My Favourite Theory.* x is permissible iff x is permissible according to the (an) imprecise decision principle that receives highest credence.

For starters, so defined, MFT can handle precise credences given to imprecise decision principles, but cannot handle every case in which imprecise credences are involved. In some such cases, there isn't any principle that can reasonably be said to have "highest credence." For example, what if in some second-order decision-situation $\mathfrak{D}^2$, $\mathbf{R}^2 \in \mathfrak{D}^2$ is such that Maximality is assigned highest credence relative to $\mathsf{C}_1 \in \mathbf{R}^2$, while Liberal is assigned highest credence relative to $C_2 \in \mathbf{R}^2$? Neither Maximality, nor Liberal, can reasonably be said to have "highest credence" without reference to a particular credence function. This means that, unfortunately, MFT cannot be well-defined as a second-order decision principle. So, while I've repeatedly called MFT a second-order decision principle, this claim was, in fact, incorrect in so far as we want to be able to precisely define second-order decision principles.

That said, it's possible to use MFT as a stepping stone for various full-blown and well-defined second-order decision principles. Basically, this is achieved by letting MFT take over MEV's role in the imprecise decision principles of the sort discussed in §§2.4-2.5, i.e., principles that assess alternatives on the basis of their expected value relative to some (not necessarily all) credence functions in the representor. Let me explain.

Recall that MEV is designed for cases in which precise credences are assigned to outcomes that might be brought about (see §2.3). Nevertheless, if imprecise credences are assigned to outcomes, it's still possible to talk about the expected value of alternatives–so long as we're talking about expected value relative to credence functions in the representor. Keeping this in mind, we can use MEV as a stepping stone for the development of imprecise decision principles such as

> *Liberal (informal).* x is permissible iff x maximizes expected value relative to at least one credence function in the representor; otherwise, x is impermissible.

Similarly, MFT is designed only for cases in which precise credences are assigned to imprecise decision principles. Nevertheless, in the imprecise credence case, it remains possible to talk about "most plausible" imprecise decision principles or principles that receive "highest credence"–so long as we're talking about "most plausible" or receiving "highest credence" relative to credence functions in the representor. Keeping this in mind, we can use MFT as a stepping stone for the development of second-order decision principles such as

> *Liberal-My Favourite Theory (informal).* x is permissible iff x is permissible according to the (an) imprecise decision principle that is assigned highest credence relative to at least one credence function in the representor; otherwise, x is impermissible.

The difference between Liberal and Liberal-My Favourite Theory (Liberal-MFT) is essentially that the former is concerned with the expected values of alternatives relative to the credence functions in the representor, and the latter with the plausibility of imprecise decision principles relative to the credence functions in the representor.

Using subscript LIB-MFT to designate Liberal-MFT's choice- and prohibited-sets, we can state this second-order decision principle formally as

> *Liberal-My Favourite Theory (formal).* For any $\mathfrak{D}^2$, let $x \in \mathbf{A} \in \mathfrak{D}^2$ be an element of $CS_{LIB\text{-}MFT}$ iff there exists at least one $C_i \in \mathbf{R}^2 \in \mathfrak{D}^2$ such that x is an element of CS_{P1} according to some $P^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$ and there exists no $P^{1\prime} \in \mathbf{P}^1 \in \mathfrak{D}^2$ for which $C_i(P^{1\prime}) > C_i(P^1)$; otherwise, x is an element of $PS_{LIB\text{-}MFT}$.

To illustrate Liberal-MFT, let's apply it to the running example, i.e., Nina's case (from §§2.4-2.5; 3.2). Recall that Nina faces a decision between making a €100 donation to either the Against Malaria Foundation (x), Animal Charity Evaluators (y), or the Patient Philanthropy Fund (z). The only imprecise decision principles to which she gives credence are Liberal, Moderate, and Maximality. According to Liberal, x and y are permissible and z is impermissible (i.e., $CS_{LIB}=\{x, y\}$ and $PS_{LIB}=\{z\}$); according to Moderate, x and y are indeterminately permissible and z is impermissible (i.e., $CS_{MOD}=\{\}$ and $PS_{MOD}=\{z\}$); according to Maximality, x, y, and z are permissible (i.e., $CS_{MAX}=\{x, y, z\}$ and $PS_{MAX}=\{\}$). Nina's credences with respect to these principles are captured in credence functions C_3 and C_4. Credence function C_3 is such that $C_3(\text{Liberal})=0.5$; $C_3(\text{Moderate})=0.3$; $C_3(\text{Maximality})=0.3$. Credence function C_4 is such that $C_4(\text{Liberal})=0.2$; $C_4(\text{Moderate})=0.5$; $C_4(\text{Maximality})=0.3$.

Liberal-MFT has this to say: Since x and y are permissible according to Liberal, which receives highest credence relative to C_3, they are permissible. Since z isn't permissible according to any imprecise decision principle that receives highest credence relative to any of the credence functions (Liberal, Moderate), it's impermissible. Formally, since $C_3 \in \mathbf{R}^2 \in \mathfrak{D}^2$ is such that x is an element of CS_{LIB} according to Liberal $\in \mathbf{P}^1 \in \mathfrak{D}^2$ and there exists no $P^{1\prime} \in \mathbf{P}^1 \in \mathfrak{D}^2$ for which $C_3(P^{1\prime}) > C_3(\text{Liberal})$, $x \in CS_{LIB\text{-}MFT}$. Since $C_3 \in \mathbf{R}^2 \in \mathfrak{D}^2$ is such that y is an element of CS_{LIB} according to Liberal $\in \mathbf{P}^1 \in \mathfrak{D}^2$ and there exists no $P^{1\prime} \in \mathbf{P}^1 \in \mathfrak{D}^2$ for which $C_3(P^{1\prime}) > C_3(\text{Liberal})$, $y \in CS_{LIB\text{-}MFT}$. Since there's no $C_i \in \mathbf{R}^2 \in \mathfrak{D}^2$ such that z is an element of CS_{P1} according to some $P^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$ and there exists no $P^{1\prime} \in \mathbf{P}^1 \in \mathfrak{D}^2$ for which $C_i(P^{1\prime}) > C_i(P^1)$, $z \in PS_{LIB\text{-}MFT}$. In sum, $CS_{LIB\text{-}MFT}=\{x, y\}$ and $PS_{LIB\text{-}MFT}=\{z\}$.

Notably, all imprecise decision principles of the sort discussed in §§2.4-2.5 have the property of coinciding with MEV in case precise credences are assigned to outcomes conditional on choosing the alternatives that might bring them about. Liberal-MFT seems to inherit this property, in a way: If precise credences are (in effect) assigned to imprecise decision principles–i.e., for any $\mathfrak{D}^2$ such that for every pair C_i, $C_j \in \mathbf{R}^2 \in \mathfrak{D}^2$ and every $P^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$, $C_i(P^1)=C_j(P^1)$–Liberal-MFT coincides with MFT. Hence Liberal-MFT can handle precise as well as imprecise credences given to imprecise decision principles.

We can rework other imprecise decision principles in a similar manner,[16,17] like so

> *Moderate-My Favourite Theory (informal).* x is permissible iff relative to every credence function in the representor, x is permissible according to the (an) imprecise decision principle that is assigned highest credence; x is impermissible iff relative to every credence function, x is impermissible according the (an) imprecise decision principle that is assigned highest credence; otherwise, x is indeterminately permissible.
>
> *Maximality-My Favourite Theory (informal).* x is permissible iff there is no y such that y is strictly preferred to x; otherwise, x is impermissible. Furthermore, x is strictly preferred to y iff relative to every credence function in the representor, x should be chosen over y according to the (an) imprecise decision principle that is assigned highest credence; x is indifferent to y iff relative to every credence function, the (an) imprecise decision principle that is assigned highest credence is indifferent whether x or y is chosen; otherwise, preferences between x and y are indeterminate.

Using subscript MO-MFT and subscript MA-MFT to designate Moderate-My Favourite Theory's (Moderate-MFT's) and Maximality-My Favourite Theory's (Maximality-MFT's) choice- and prohibited-sets, respectively, we can state them formally as

---

[16] I'm not sure whether *every* imprecise decision principle of the sort discussed in §§2.4-2.5 can be reworked as a second-order decision principle in this manner. Fortunately, for present purposes, nothing hangs on this.

[17] I'm not claiming that the way in which I combine imprecise decision principles with second-order decision principles from the literature on moral uncertainty is the only way to go about this, or the best way (which presumably involves a formalization of 'choice-worthiness'; see §3.1 and §4.2.3). I merely claim that it's a possible way of going about it, one that leads to well-defined second-order decision principles for decision-making under uncertainty about imprecise decision principles.

Moderate-My Favourite Theory (formal). For any $\mathfrak{D}^2$, let $x \in \mathbf{A} \in \mathfrak{D}^2$ be an element of $CS_{MOD\text{-}MFT}$ iff every $C_i \in \mathbf{R}^2 \in \mathfrak{D}^2$ is such that x is an element of CS_{P1} according to some $P^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$ and there exists no $P^{1\prime} \in \mathbf{P}^1 \in \mathfrak{D}^2$ for which $C_i(P^{1\prime})>C_i(P^1)$; let x be an element of $PS_{MOD\text{-}MFT}$ iff every C_i is such that x is an element of PS_{P1} according to some P^1 for which there exists no $P^{1\prime}$ such that $C_i(P^{1\prime})>C_i(P^1)$.

Maximality-My Favourite Theory (formal). For any $\mathfrak{D}^2$, let $x \in \mathbf{A} \in \mathfrak{D}^2$ be an element of $CS_{MAX\text{-}MFT}$ iff there is no $y \in \mathbf{A} \in \mathfrak{D}^2$ such that y is strictly preferred to x; otherwise, x is an element of $PS_{MAX\text{-}MFT}$. Furthermore, x is strictly preferred to y iff relative to every $C_i \in \mathbf{R}^2 \in \mathfrak{D}^2$, according to some P^1 for which there exists no $P^{1\prime}$ such that $C_i(P^{1\prime})>C_i(P^1)$, if $x \in CS_{P1}$, then $y \notin CS_{P1}$, and if $x \notin CS_{P1}$, then $y \in PS_{P1}$; x is indifferent to y iff relative to every C_i, according to some P^1 for which there exists no $P^{1\prime}$ such that $C_i(P^{1\prime})>C_i(P^1)$, $x \in CS_{P1}$ iff $y \in CS_{P1}$, and $x \in PS_{P1}$ iff $y \in PS_{P1}$; otherwise, preferences between x and y are indeterminate.

To clarify these second-order decision principles, let's apply them to Nina's case as outlined just above.

Moderate-MFT has this to say: x and y are indeterminately permissible since they are permissible according to Liberal, which receives highest credence relative to C_3, but not permissible according to Moderate, which receives highest credence relative to C_4. z is impermissible, since relative to both credence functions, z is impermissible according the (an) imprecise decision principle that is assigned highest credence (i.e., Liberal, Moderate). Formally, since $C_3 \in \mathbf{R}^2 \in \mathfrak{D}^2$ is such that x and y are elements of CS_{LIB} according to Liberal $\in \mathbf{P}^1 \in \mathfrak{D}^2$ and there exists no $P^{1\prime} \in \mathbf{P}^1 \in \mathfrak{D}^2$ for which $C_3(P^{1\prime})>C_3(\text{Liberal})$, $x, y \notin PS_{MOD\text{-}MFT}$. Since $C_4 \in \mathbf{R}^2 \in \mathfrak{D}^2$ is such that x and y aren't elements of CS_{MOD} according to Moderate $\in \mathbf{P}^1 \in \mathfrak{D}^2$ and there exists no $P^{1\prime} \in \mathbf{P}^1 \in \mathfrak{D}^2$ for which $C_4(P^{1\prime})>C_4(\text{Moderate})$, $x, y \notin CS_{MOD\text{-}MFT}$. Since every C_i is such that z is an element of PS_{P1} according to some P^1 for which there exists no $P^{1\prime}$ such that $C_i(P^{1\prime})>C_i(P^1)$ (i.e., Liberal and Moderate), $z \in PS_{MOD\text{-}MFT}$. In sum, $CS_{MOD\text{-}MFT}=\{\}$ and $PS_{MOD\text{-}MFT}=\{z\}$.

Maximality-MFT has this to say: x and y should be strictly preferred to z, since relative to C_3 as well as C_4, the imprecise decision principle that receives highest credence (i.e., Liberal, Moderate) wants x and y to be chosen over z. Other preferences between the alternatives are indeterminate. Hence only z is such that there's at least one alternative strictly preferred to it, making z impermissible and x and y permissible. Formally, x and y should be strictly preferred to z, since (a) relative to $C_3 \in \mathbf{R}^2 \in \mathfrak{D}^2$, according to Liberal for which there exists no $P^{1\prime} \in \mathbf{P}^1 \in \mathfrak{D}^2$ such that $C_3(P^{1\prime})>C_3(\text{Liberal})$, $x, y \in CS_{LIB}$ and $z \notin CS_{LIB}$; and (b) relative to $C_3 \in \mathbf{R}^2 \in \mathfrak{D}^2$, according to Moderate for which there exists no $P^{1\prime}$ such that $C_3(P^{1\prime})>C_3(\text{Moderate})$, $x, y \notin CS_{MOD}$ and $z \in PS_{MOD}$. Since other preferences between the alternatives are indeterminate, it follows that $CS_{MAX\text{-}MFT}=\{x, y\}$ and $PS_{MAX\text{-}MFT}=\{z\}$.

So, MFT gives rise to at least three second-order decision principles: Liberal-MFT, Moderate-MFT, and Maximality-MFT. I'll now discuss the second-order decision principles presented in response to the objections to MFT, in the order as given.

Just as MFT,

> *My Favourite Alternative.* x is permissible iff x is the (an) alternative that is most likely permissible.

isn't a proper second-order decision principle because it isn't suited for cases where imprecise credences are given to imprecise decision principles. An alternative x can be most likely permissible relative to credence function C_1, while an alternative y can be most likely permissible relative to credence function C_2.

Just as MFT, My Favourite Alternative (MFA) can be used as a stepping stone for various full-blown and well-defined second-order decision principles, by letting MFA take over MEV's role in the imprecise decision principles, for example, as follows

> *Liberal-My Favourite Alternative (informal).* x is permissible iff x is most likely permissible relative to at least one credence function in the representor; otherwise, x is impermissible.

Using subscript LIB-MFA to designate Liberal-My Favourite Alternative's (Liberal-MFA's) choice- and prohibited-sets, we can state this second-order decision principle formally as

> *Liberal-My Favourite Alternative (formal).* For any $\mathfrak{D}^2$, let $x \in \mathbf{A} \in \mathfrak{D}^2$ be an element of $CS_{LIB\text{-}MFA}$ iff there exists at least one $C_i \in \mathbf{R}^2 \in \mathfrak{D}^2$ such that x is most likely an element of CS_R relative to C_i; otherwise, x is an element of $PS_{LIB\text{-}MFA}$.

By now you'll hopefully get the trick, so allow me to skip the illustration and put the other newly developed second-order decision principles in a footnote.[18]

With regards to Dominance-My Favourite Theory (Dominance-MFT), for our purposes it can be redefined as

> *Dominance-My Favourite Theory.* x is permissible iff x is permissible according to the (an) imprecise decision principle that is most plausible, such that x is not dominated.

Again, Dominance-MFT doesn't work for every imprecise credence case, but does if conjoined with imprecise decision principles of the sort discussed in §§2.4-2.5, as I demonstrate in this footnote.[19]

---

[18] *Moderate-My Favourite Alternative (informal).* x is permissible iff relative to every credence function in the representor, x is most likely permissible; x is impermissible iff relative to every credence function, x is most likely impermissible; otherwise, x is indeterminately permissible.

Maximality-My Favourite Alternative (informal). x is permissible iff there is no y such that y is strictly preferred to x; otherwise, x is impermissible. Furthermore, x is strictly preferred to y iff x is more likely to be permissible than y relative to every credence function in the representor; x is indifferent to y iff x is as likely to be permissible as y relative to every credence function; otherwise, preferences between x and y are indeterminate.

Using subscript MOD-MFA and subscript MAX-MFA to designate Moderate-My Favourite Alternative's (Moderate-MFA's) and Maximality-My Favourite Alternative's (Maximality-MFA's) choice- and prohibited-sets, we can state these second-order decision principles formally as

Moderate-My Favourite Alternative (formal). For any $\mathfrak{D}^2$, let $x \in \mathbf{A} \in \mathfrak{D}^2$ be an element of $CS_{\text{MOD-MFA}}$ iff every $C_i \in \mathbf{R}^2 \in \mathfrak{D}^2$ is such that x is most likely an element of CS_R relative to C_i; let x be an element of $PS_{\text{MOD-MFA}}$ iff every C_i is such that x is most likely an element of PS_R relative to C_i.

Maximality-My Favourite Alternative (formal). For any $\mathfrak{D}^2$, let $x \in \mathbf{A} \in \mathfrak{D}^2$ be an element of $CS_{\text{MAX-MFA}}$ iff there is no $y \in \mathbf{A} \in \mathfrak{D}^2$ such that y is strictly preferred to x; otherwise, x is an element of $PS_{\text{MAX-MFA}}$. Furthermore, x is strictly preferred to y iff relative to every $C_i \in \mathbf{R}^2 \in \mathfrak{D}^2$, x is more likely to be an element of CS_R than y; x is indifferent to y iff relative to every C_i, x is as likely to be an element of CS_R as y; otherwise, preferences between x and y are indeterminate.

[19] *Liberal-Dominance-My Favourite Theory (informal).* x is permissible iff x is permissible according to the (an) imprecise decision principle that is assigned highest credence, such that x is not dominated, relative to at least one credence function in the representor; otherwise, x is impermissible.

Moderate-Dominance-My Favourite Theory (informal). x is permissible iff relative to every credence function in the representor, x is permissible according to the (an) imprecise decision principle that is assigned highest credence, such that x is not dominated; x is impermissible iff relative to every credence function, x is impermissible according the (an) imprecise decision principle that is assigned highest credence or x is dominated; otherwise, x is indeterminately permissible.

Maximality-Dominance-My Favourite Theory (informal). x is permissible iff there is no y such that y is strictly preferred to x; otherwise, x is impermissible. Furthermore, x is strictly preferred to y iff relative to every credence function in the representor, x should be chosen over y according to the (an) imprecise

With regards to

> *Maximize Expected Choice-Worthiness (formal).* x is permissible iff there's no available alternative y such that $\mathrm{EC}(y)>\mathrm{EC}(x)$.

matters are more complicated. For MEC to serve as a second-order decision principle for the sake of decision-making in the face of uncertainty about imprecise decision principles, the expected choice-worthiness EC of available alternatives has to be well-defined. Given the definition of an alternative's EC, this in turn presupposes, firstly, that choice-worthiness is intertheoretically comparable across imprecise decision principles. Tarsney's (2018) approach referenced in §4.2.4 doesn't work for this purpose, but MacAskill (2016b, fn. 13) has in passing suggested an approach for making intertheoretic comparisons of choice-worthiness

---

decision principle that is assigned highest credence, such that x isn't dominated by y; x is indifferent to y iff relative to every credence function, the (an) imprecise decision principle that is assigned highest credence is indifferent whether x or y is chosen; otherwise, preferences between x and y are indeterminate.

Using subscript LIB-D-MFT, subscript MOD-D-MFT, and subscript MAX-D-MFT to designate Liberal-Dominance-My Favourite Theory's (Liberal-D-MFT's), Moderate-Dominance-My Favourite Theory's (Moderate-D-MFT's), and Maximality-Dominance-My Favourite Theory's (Maximality-D-MFT's) choice- and prohibited-sets, respectively, we can state these second-order decision principles formally as

Liberal-Dominance-My Favourite Theory (formal). For any $\mathfrak{D}^2$, let $x \in \mathbf{A} \in \mathfrak{D}^2$ be an element of $\mathrm{CS}_{\text{LIB-D-MFT}}$ iff there exists at least one $\mathrm{C_i} \in \mathbf{R}^2 \in \mathfrak{D}^2$ such that (i) x is an element of $\mathrm{CS}_{\mathrm{P1}}$ according to some $\mathrm{P}^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$ and there exists no $\mathrm{P}^{1\prime} \in \mathbf{P}^1 \in \mathfrak{D}^2$ for which $\mathrm{C_i}(\mathrm{P}^{1\prime})>\mathrm{C_i}(\mathrm{P}^1)$, and (ii) there is no y such that $y \in \mathrm{CS}_{\mathrm{P1}}$ according to every $\mathrm{P}^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$, while $x \in \mathrm{PS}_{\mathrm{P1}}$ according to some $\mathrm{P}^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$; otherwise, x is an element of $\mathrm{PS}_{\text{LIB-D-MFT}}$.

Moderate-Dominance-My Favourite Theory (formal). For any $\mathfrak{D}^2$, let $x \in \mathbf{A} \in \mathfrak{D}^2$ be an element of $\mathrm{CS}_{\text{MOD-D-MFT}}$ iff every $\mathrm{C_i} \in \mathbf{R}^2 \in \mathfrak{D}^2$ is such that (i) x is an element of $\mathrm{CS}_{\mathrm{P1}}$ according to some $\mathrm{P}^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$ and there exists no $\mathrm{P}^{1\prime} \in \mathbf{P}^1 \in \mathfrak{D}^2$ for which $\mathrm{C_i}(\mathrm{P}^{1\prime})>\mathrm{C_i}(\mathrm{P}^1)$, and (i) there is no y such that $y \in \mathrm{CS}_{\mathrm{P1}}$ according to every $\mathrm{P}^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$, while $x \in \mathrm{PS}_{\mathrm{P1}}$ according to some $\mathrm{P}^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$; let x be an element of $\mathrm{PS}_{\text{MOD-D-MFT}}$ iff (i) every $\mathrm{C_i}$ is such that x is an element of $\mathrm{PS}_{\mathrm{P1}}$ according to some P^1 for which there exists no $\mathrm{P}^{1\prime}$ such that $\mathrm{C_i}(\mathrm{P}^{1\prime})>\mathrm{C_i}(\mathrm{P}^1)$ or (ii) there is some y such that $y \in \mathrm{CS}_{\mathrm{P1}}$ according to every $\mathrm{P}^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$, while $x \in \mathrm{PS}_{\mathrm{P1}}$ according to some $\mathrm{P}^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$.

Maximality-Dominance-My Favourite Theory (formal). For any $\mathfrak{D}^2$, let $x \in \mathbf{A} \in \mathfrak{D}^2$ be an element of $\mathrm{CS}_{\text{MAX-D-MFT}}$ iff there is no $y \in \mathbf{A} \in \mathfrak{D}^2$ such that y is strictly preferred to x; otherwise, x is an element of $\mathrm{PS}_{\text{MAX-D-MFT}}$. Furthermore, x is strictly preferred to y iff relative to every $\mathrm{C_i} \in \mathbf{R}^2 \in \mathfrak{D}^2$, (i) according to some P^1 for which there exists no $\mathrm{P}^{1\prime}$ such that $\mathrm{C_i}(\mathrm{P}^{1\prime})>\mathrm{C_i}(\mathrm{P}^1)$, if $x \in \mathrm{CS}_{\mathrm{P1}}$, then $y \notin \mathrm{CS}_{\mathrm{P1}}$, and if $x \notin \mathrm{CS}_{\mathrm{P1}}$, then $y \in \mathrm{PS}_{\mathrm{P1}}$, and (ii) it is not so that $y \in \mathrm{CS}_{\mathrm{P1}}$ according to every $\mathrm{P}^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$, while $x \in \mathrm{PS}_{\mathrm{P1}}$ according to some $\mathrm{P}^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$; x is indifferent to y iff relative to every $\mathrm{C_i}$, according to some P^1 for which there exists no $\mathrm{P}^{1\prime}$ such that $\mathrm{C_i}(\mathrm{P}^{1\prime})>\mathrm{C_i}(\mathrm{P}^1)$, $x \in \mathrm{CS}_{\mathrm{P1}}$ iff $y \in \mathrm{CS}_{\mathrm{P1}}$, and $x \in \mathrm{PS}_{\mathrm{P1}}$ iff $y \in \mathrm{PS}_{\mathrm{P1}}$; otherwise, preferences between x and y are indeterminate.

across causal and evidential decision theory.[20] It seems possible to employ a version of this strategy in order to make intertheoretic comparisons of choice-worthiness across imprecise decision principles, although I cannot show this here. Secondly, it presupposes that every imprecise decision principle can be associated with a choice-worthiness function CW that can be transformed into a choice-worthiness function CW' by means of a positive linear transformation. That is, choice-worthiness has to be measurable on an interval scale. Intuitively, this just means that alternatives will be ranked in terms of choice-worthiness, and we can be informed *how much more* choice-worthy some alternatives are as compared to others. Thirdly, what's more, every choice-worthiness function has to be *complete*, i.e., for every alternative x in every decision-situation $\mathfrak{D}$ and every imprecise decision principle P^1, $\mathsf{CW}_{\mathsf{P}1}(x)$ should be assigned a value. Neither of the last two requirements can be met.

With regards to the interval-scale measurability of choice-worthiness: Bales (2018, 1694) ascribes the following take on choice-worthiness to Rinard (2015)–who, recall from §2.5, originally developed and defended Moderate in her (2015): An alternative is choice-worthy if it's permissible, not choice-worthy if it's impermissible, and it's indeterminate whether it's choice-worthy if it's indeterminately permissible. If Bales's interpretation of Rinard (2015) is correct, then it seems fair to characterize Moderate as providing a ranking of alternatives in terms of choice-worthiness (with permissible alternatives declared more choice-worthy than impermissible ones), but we wouldn't do Moderate justice if we were to claim that it would tell us anything about how much more choice-worthy some alternatives are as compare to others. So, Moderate doesn't provide us with interval-scale measurability of choice-worthiness (rather, it says that choice-worthiness is *ordinally* measurable, i.e., it merely provides a ranking in terms of choice-worthiness).

With regards to the requirement that choice-worthiness be complete: If Bales's (2018, 1694) aforementioned interpretation of Rinard (2015) is correct, then Moderate stipulates that the choice-worthiness of indeterminately permissible alternatives is indeterminate. So, the choice-worthiness of alternatives according to Moderate can't always be properly captured by a complete choice-worthiness function.

[20] Causal and evidential decision theory are particular versions of MEV, but their particulars are irrelevant for present purposes. (I've been using the evidential version, but nothing of substance hangs on this).

This shows that MEC cannot be the Unique Account, since it cannot handle every case in which decision-makers have to come to a decision under uncertainty about imprecise decision principles. However, while MEC isn't able to handle the sorts of choice-worthiness functions associated with imprecise decision principles, some methods are. At risk of being charged with hand waving, I'm not going to fully demonstrate this; the following comments hopefully suffice.

The field of *social choice theory* is to some extent occupied with the development of rules for *aggregating* the *utility functions* of different individuals, i.e., combining individual utility functions ranging over certain objects into a single utility function ranging over these objects (for an introduction to social choice theory, see, e.g., Sen 2017). Formally speaking, choice-worthiness functions are essentially the same as utility functions (although one can make assumptions about the formal properties of utility functions that distinguish them from choice-worthiness functions, and *vice versa*). On a substantive level, the individual utility functions are typically understood as ranging over *social states* (i.e., complete descriptions of particular states that a society can be in), and the single utility function is then viewed as a *social* utility function. An example of an aggregation rule that makes use of this interpretation is utilitarianism, which can in the context of social choice theory be defined as saying that social state x has greater social utility than social state y iff given all individual utility functions ranging over x and y, x has greater total utility than y; x has equal social utility to y iff given all individual utility functions ranging over x and y, x and y have equal total utility (see MacAskill 2016a, 976–77).

There's nothing keeping us from reinterpreting aggregation rules developed by social choice theorists so that we can use them to aggregate the choice-worthiness functions of imprecise decision principles ranging over available alternatives into a single 'second-order' choice-worthiness function ranging over these alternatives. Otherwise put, there's nothing keeping us from reinterpreting aggregation rules from social choice theory as second-order decision principles (cf. MacAskill 2016a). (Notably, a link between imprecise decision principles and social choice theory has been drawn before (e.g., Weatherson 1998, 6; R. C. Bradley 2017, 267; Mogensen and Thorstad 2020, 21). The different credence functions in the representor have been treated as different individuals, each with their own 'utilities' pertaining to the available alternatives. Conceived of as such, imprecise decision principles are aggregation rules).

That said, an important caveat is that not *all* aggregation rules can be reasonably reemployed as second-order decision principles. This is because, firstly, some aggregation rules incorporate information about intertheoretic (interpersonal) comparability of choice-worthiness (utility) across imprecise decision principles (individuals), and some do not. Whether it's indeed possible to make such comparisons will have an impact on which sort of aggregation rule we'll want to use. As repeatedly noted, I think it's indeed possible to make such comparisons, but this is an issue that warrants further investigation.

Secondly, some aggregation rules cannot accommodate the types of choice-worthiness functions associated with imprecise decision principles, such as Moderate's incomplete choice-worthiness function that measures choice-worthiness on an ordinal scale. Incomplete functions aren't always admissible as input for aggregation rules, and neither are functions that measure choice-worthiness (utility) on an ordinal scale.

Thirdly, in cases of rational uncertainty about imprecise decision principles, decision-makers give credence to various imprecise decision principles. These credences need not be equal. It's possible (likely, even) that some imprecise decision principles are considered more plausible than others. This means that for the purpose of aggregation, the choice-worthiness functions of the imprecise decision principles plausibly have to be *weighted*. Aggregation rules that give weights have been developed, so this hurdle can be overcome (MacAskill 2016a has developed and defended a 'Credence-Weighted Borda Rule'). But, of course, I'm assuming that it can be rational for decision-makers to give imprecise credences to imprecise decision principles. To my knowledge, the typical aggregation rule that gives weights to utility functions, assumes that weights are precise (MacAskill 2016a, fn. 52 is an exception, suggesting a version of his 'Credence-Weighted Borda Rule' that can handle imprecise credences, which bears some resemblance to Moderate and Maximality). Therefore, these rules are inapplicable. But this is fixable: We can use the same trick as with MFT, MFA, and Dominance-MFT. That is, we can take an aggregation rule, and let it take over MEV's role in imprecise decision principles of the sort discussed in §§2.4-2.5 (e.g., we might get a 'Moderate-Credence-Weighted Borda Rule').

All this should be fleshed out further, but space and time don't permit me to delve into this. I take it that I've made it sufficiently plausible that a whole class of second-order decision principles can be borrowed or adapted from the literature on social choice theory.

So, in sum, while the Metanormativist might tell us that we should employ a single second-order decision principle to take uncertainty about imprecise decision principles into account in our decision-making, we now seem to be spoilt for choice: I've given nine examples of second-order decision principles and pointed the reader to even more. Each is a hybrid of (i) a second-order decision principle from the literature on moral uncertainty–or the literature on social choice theory–and (ii) an imprecise decision principle of the sort discussed in §§2.4-2.5. The Metanormativist who endorses a Unique Account, should say that one of these principles, or some alternative, should be privileged over all others. Next, I'll argue that there's no second-order decision principle that can plausibly play this part.

§4.4 Against Privileging Any Second-Order Decision Principle

In this subsection, I'll argue that there isn't any one second-order decision principle for which it can plausibly be claimed that decision-makers should always follow its prescriptions, i.e., that there's no plausible Unique Account. I'll focus on cases in which decision-makers are rationally *uncertain* about second-order decision principles to prove the point. Clearly such cases exist. Recall that, at the end of §2.5, I remarked that it wasn't clear which imprecise decision principle, if any, effective altruists ought to conform to in the face of cluelessness. If this isn't due to irrationality, it must be possible to be rationally uncertain about which imprecise decision principle to adopt. It similarly isn't clear which second-order decision principle, if any, to adopt. I take it that this too doesn't show us to be irrational. Moreover, on the assumption that it's possible to be rationally uncertain about imprecise decision principles (as Metanormativists would claim), it seems impossible to reasonably deny that it can be rational to be uncertain about second-order decision principles, considering that these are *variations on imprecise decision principles.* To be fair, the imprecise decision principles differ in important respects from the second-order decision principles based on them. For instance, while Liberal is concerned with the expected values of alternatives relative to the credence functions in the representor, Liberal-MFT is occupied with the plausibility of imprecise decision principles relative to the credence functions in the representor. Nevertheless, there are important similarities that shouldn't be overlooked. For example, Liberal and Liberal-MFT agree that the 'verdict' of a single credence function in the representor suffices for an alternative to count as permissible. Given that we're uncertain about Liberal, it would be unreasonable of us to be fully certain about Liberal-MFT (i.e. to give it credence 1 on every credence function in our representor) or to reject it out right (i.e., to give it credence 0 on every credence function).

This means that the proponent of any Unique Account must claim that some second-order decision principle should be privileged over all others, in spite of the fact that decision-makers can–or, moreover, should–be rationally uncertain about it. This is only reasonably if there's a suitable candidate for this role. There isn't.

The most important reason for the non-existence of such a second-order decision principle is just that all such principles are variations on imprecise decision principles. It hardly makes senses to claim that, for instance, uncertainty about Liberal should be accounted for, but that uncertainty about Liberal-MFT shouldn't, and that decision-makers who are uncertain about imprecise decision principles such as Liberal should necessarily deal with this uncertainty in their decision-making using Liberal-MFT.

The proponent of a Unique Account might retort that this argument references the second-order decision principles that I've developed. In fact, it depends crucially on them in so far as I'm assuming that *all* second-order decision principles are variations on imprecise decision principles. She might hold that there exists some second-order decision principle that isn't based on any imprecise decision principle, and that this is the one that decision-makers should always obey.

My reply is this: Perhaps such a second-order decision principle exists, but I'm not so sure–I need to see it to believe it. And the burden of proof rests with the Metanormativist.

Three more (related) arguments converge on the conclusion that there's no single second-order decision principle that decision-makers should always act in accordance with for the purpose of dealing with uncertainty about imprecise decision principles. I'll present them in order of increasing strength. First, each second-order decision principle covered in the previous subsection can be subjected to criticism. To give just a few examples, the second-order decision principles based on MFT, MFA, and Dominance-MFT fail to take intertheoretic comparisons of choice-worthiness across imprecise decision principles into account. Intuitively, second-order decision principles seem required to take these into consideration, at least in some cases (granted that such comparisons can indeed be made, across at least some imprecise decision principles). With regards to any aggregation rule borrowed from the literature on social choice theory, note that numerous *impossibility results* have been proved for aggregation rules (see, e.g., Sen 2017). Basically, these results show that any aggregation rule will violate at least one intuitively attractive condition (e.g., some violate Sen's Condition β; see §4.2). We can plausibly suppose that the aggregation rules will not escape from these impossibility results if we treat them as second-order decision principles.

Second, proponents of Unique Accounts owe us an *explanation* of why a single second-order decision principle should always be used, even in the face of uncertainty about it. But such an explanation doesn't seem to be forthcoming.

Finally, the arguments in favour of Metanormativism sketched in §3.1 work equally well to defend the claim that rational uncertainty about second-order decision principles should rationally be accounted for in decision-making. This claim has *intuitive appeal*; it's *arbitrary* to take uncertainty about imprecise decision principles into consideration, but to ignore uncertainty about second-order decision principles; second-order decision principles about which we're uncertain don't seem sufficiently *action guiding*, for how can a view of which we don't know that we ought to follow it help us identify which actions to perform? To the extent that Metanormativists find these arguments compelling evidence for their own position, they should find them compelling evidence for the position that rational uncertainty about second-order decision principles should rationally be accounted for in decision-making.

To round off, let's discuss Tarsney's (2019) Unique Account. At this stage, we'll have all the tools to show that it's implausible, at least in the present context. Tarsney argues that decision-makers should always obey the second-order decision principle labelled *the Enkratic Principle* in the domain of decision-making under normative uncertainty. Even though this isn't exactly what Tarsney had in mind, I'll treat it as a second-order decision principle for the purpose of dealing with uncertainty about imprecise decision principles. So, while I'll reject the Enkratic Principle when conceived of as such, I'm not rejecting it wholesale (e.g., I'm not excluding that decision-makers should always conform to the Enkratic Principle when making decisions under moral or prudential uncertainty).

In our terminology, the principle can, as a first approximation, be defined as

> *The Enkratic Principle.* If a decision-maker believes that she should objectively choose x, then she should rationally choose x (see Tarsney 2019, 21).

So defined, as Tarsney points out, the Enkratic Principle cannot deal with uncertainty in a plausible way. Suppose that a decision-maker believes that she should objectively choose x, but isn't certain that she should. In particular, suppose that she gives some (though not much) credence to the proposition that she should objectively choose y rather than x. The reasons for choosing y are very strong, and those for choosing x very weak. If so, it seems more plausible to say that she should rationally choose y than that she should rationally choose x–contra the Enkratic Principle as it's now formulated. But perhaps belief should entail certainty. If so, the Enkratic Principle ("If a decision-maker is *certain* that she should objectively choose x, then she should rationally choose x) doesn't apply very often. Tarsney directs our attention to Wedgwood (2013), who has proposed a version of the Enkratic Principle that is suited to deal with uncertainty. Setting aside irrelevant details, and tailored to our terminology and purposes, it's a second-order decision principle that we've seen in §4.2.3 and §4.3, namely

> *Maximize Expected Choice-Worthiness (formal).* x is permissible iff there's no available alternative y such that $\mathsf{EC}(y)>\mathsf{EC}(x)$.

To briefly rehears the arguments for why MEC was inadequate as a second-order decision principle for decision-making in the face of uncertainty about imprecise decision principles: The expected choice-worthiness EC of available alternatives has to be well-defined. This requires that choice-worthiness is intertheoretically comparable across imprecise decision principles (which I think is possible), as well as that every imprecise decision principle can be associated with a *complete* choice-worthiness function CW that measures choice-worthiness on an *interval scale* (which I've shown to be impossible).

But, for argument's sake, assume that these problems can be remedied. Another issue immediately arises: Just as MEV, MEC cannot deal with imprecise credences. Similar to the expected value EV of alternatives, the expected choice-worthiness EC of any alternative is defined relative to a single credence function.

While Tarsney doesn't address the possibility of imprecise credence, Wedgwood (2013, 495–496) proposes a necessary and sufficient condition for $\mathsf{EC}(y)>\mathsf{EC}(x)$ that can accommodate them. Again setting aside irrelevant details, and again tailored to our terminology and purposes, Wedgwood suggests that $\mathsf{EC}(y)>\mathsf{EC}(x)$ iff relative to every credence function in the representor, $\mathsf{EC}(y)>\mathsf{EC}(x)$. (This definition of '$\mathsf{EC}(y)>\mathsf{EC}(x)$' is poorly formulated: The *ex-*

planandum is part of the *explanans*, which is a no-go. Wedgwood doesn't make this same mistake, but working around it would make things needlessly complicated). What's presently crucial about Wedgwood's suggested necessary and sufficient condition for $\mathrm{EC}(y) > \mathrm{EC}(x)$ is that it bears a strong resemblance to imprecise decision principles such as Moderate and Maximality. It's similar to Moderate's necessary and sufficient condition for the permissibility of x, namely that x maximizes expected value relative to every credence function in the representor. It's also similar to Maximality's necessary and sufficient condition for strict preference of x over y, namely that x has greater expected value than y relative to all credence functions in the representor. Wedgwood's proposal differs markedly from Liberal, however. To mimic Liberal, it'd have to be something such as: $\mathrm{EC}(y) > \mathrm{EC}(x)$ iff relative *to at least one* credence function in the representor, $\mathrm{EC}(y) > \mathrm{EC}(x)$. So, the version of MEC that can accommodate imprecise credences falls prey to the same objection as the second-order decision principles discussed in the earlier in this subsection: The proponent of this version of MEC as the Unique Account must claim that uncertainty about imprecise decision principles should be accounted for in decision-making, while maintaining that uncertainty about a second-order decision principle that's like some imprecise decision principles, and very much unlike others, shouldn't be accounted for in decision-making. This is implausible.

§4.5 Conclusions about Unique Accounts

So, I conclude, and so should the Metanormativist, that there's no adequate Unique Account, i.e., there's not a unique second-order decision principle that should be put to work to factor in uncertainty about imprecise decision principles in every second-order decision-situation.

The Metanormativist might admit that no such principle exists. Instead, she might claim that uncertainty about second-order decision principles should be accounted for in decision-making. More precisely, that it should be dealt with using a third-order decision principle. She might acknowledge that this move will lead to an infinite regress, but deny that it tells against her position. I'll argue that it does.

§5. Against Hierarchy Accounts

§5.1 Set Up

In this section, I'll argue that there's no adequate Hierarchy Account, i.e., no adequate Metanormativist account of rational choice that is comprised of an infinite hierarchy of second- and higher order decision principles to which decision-makers give credence. Positions that there *are* adequate Hierarchy Accounts of sorts (albeit in the precise credence case) have been endorsed by Sepielli (2014), MacAskill (2014, 217–219), and Trammell (2021). Given the failure of Unique Accounts, the success of the argument against this position entails–in combination with the claim from §3.1 that Unique Accounts and Hierarchy Accounts are the only serious contenders–that Metanormativism is false.

The argument against Hierarchy Accounts will, unsurprisingly, take the form of an *infinite regress argument*: Infinite regress arguments have been developed and discussed before in the literature on decision-making under normative uncertainty and, more narrowly, on moral, prudential, or decision-theoretic uncertainty (see besides the authors just referenced, e.g., Weatherson 2014, 155–57; 2019, 13–17; Tarsney 2019; MacAskill, Bykvist, and Ord 2020, 30–33; Riedener 2021, 16–18). It's beyond the scope of this thesis to rehearse these arguments, and, so, instead, whenever I make a move that goes contrary to a significant claim from the literature, I'll make this explicit and defend making the move in spite of the disagreement (with one notable exception: I'll not engage properly with Sepielli 2014, as he invokes different senses of the concept of 'rationality.' As noted in §2.1, actions are commonly perceived as rational (irrational) iff, roughly, they (don't) cohere with our beliefs about our aims. Other than this, I'm not going to say anything about the meaning of 'rationality').

Painting in broad strokes, the infinite regress argument runs as follows. I'll focus on one particular (somewhat abstract) case in which infinite regress arises. By itself, this doesn't say much. The Metanormativist might argue it's harmless, and that rational decision-making isn't (necessarily) impaired. For example, consider a decision-maker who gives credence to only to imprecise, second- and higher order decision principles that all declare only x permissible. In this case, the decision-maker will be rationally certain that x is the only permissible alternative, since she rationally gives credence only to principles that converge on this point. Here, the Metanormativist might plausibly (at least on first sight) hold that she ought to choose x. Claims of this sort have been made by MacAskill (2014, 218–19), Sepielli (2017, 114–15), and

Trammell (2021) (see also Tarsney 2019, fn. 11). However, in the case of regress on which I'll focus, no convergence occurs. At every order, the decision principles to which the decision-maker gives credence disagree about which alternatives are permissible. In §5.2, I'll spell out conditions that lead to such regress–conditions that any remotely plausible Hierarchy Account should respect. In §5.3, I'll argue for the case under consideration that, according to any remotely plausible Hierarchy Account, it's left unspecified whether any alternative is rationally permissible. As I'll point out in §5.4, this violates

> *Decisiveness (informal).* In any decision-situation, at least one available alternative is permissible.

However, I'll stipulate that the decision-maker gives credence *only* to decision principles that meet Decisiveness, at every order. So, this decision-maker can be rationally certain that a decision principle that meets Decisiveness is true and, so, certain that Decisiveness is true. Nevertheless, if what she should rationally do is responsive to her uncertainty about decision principles–in her case, a state of rational uncertainty that entails that Decisiveness must be true–, then Decisiveness must be false. This implication is unacceptable, for two reasons.

First, it cannot be rational to be rationally certain that Decisiveness is true and, at the same time, give credence to the proposition that Decisiveness is false, which the decision-maker might if she were to embrace the Hierarchy Account. Second, the proponent of any Hierarchy Account should claim that whether a Hierarchy Account entails a condition that isn't universally met by imprecise decision principles (such as Decisiveness), depends on whether this condition is met by the decision principles that jointly form the Account. Hence, the infinite regress is *vicious*, i.e., there's no plausible Hierarchy Account. In §5.5, I'll respond to objections on behalf of the Metanormativist.

§5.2 Infinite Regress Arises

In this subsection, I'll argue that any remotely plausible Hierarchy Account should accept that there's a possible case in which the following seven conditions are met. None of them strike me particularly controversial for the Metanormativist. Jointly, the conditions entail that an infinite regress arises, such that at every order, the decision principles to which the decision-maker gives credence disagree about the permissibility of the alternatives (as I'll show at the end of this subsection).

Condition (i). Metanormativism is true.

Condition (ii). A first-order decision-situation $\mathfrak{D}$ is such that (a) imprecise decision principle P_i^1 says that only x is permissible, and all other available alternatives are impermissible; (b) imprecise decision principle P_j^1 says that only y is permissible, and all other available alternatives are impermissible.[21]

Condition (iii). P_i^1 and P_j^1 are the only imprecise decision principles to which the decision-maker gives credence.

To be sure, condition (iii) doesn't entail that the decision-maker is certain that all other imprecise decision principles are false. She merely needs to withhold judgment about them. This is reasonable to suppose if she's, for instance, simply never entertained any imprecise decision principles apart from P_i^1 and P_j^1.

Condition (iv). P_i^2 and P_j^2 are the only second-order decision principles to which the decision-maker gives credence, and the decision-maker is aware of the fact that each can be naturally reworked as n^{th}-order decision principles for any $n>2$.

Just as with condition (iii), we can assume that second-order decision principles P_i^2 and P_j^2 are the only ones that the decision-maker has ever entertained. Regarding the possibility of naturally reworking the principles as n^{th}-order decision principles for any $n>2$: This might not be possible for every possible second-order decision principle, but each second-order decision principle presented in §4.3 meets this requirement. I'm not going to show this for each of them, but I'll show for a couple how we can go about transforming them into n^{th}-order decision principles for any $n>2$.

Recall the second-order decision principle dubbed

Liberal-My Favourite Theory (informal). x is permissible iff x is permissible according to the (an) imprecise decision principle that is assigned highest credence relative to at least one credence function in the representor; otherwise, x is impermissible.

[21] The imprecise decision principles that I've discussed don't lend themselves particularly well for the roles of P_i^1 and P_j^1 (for various reasons), hence I'm keeping it abstract. Imprecise decision principles that could play the part do exist, of course; see some of the principles discussed in the literature referenced in §2.4.

Liberal-My Favourite Theory (formal). For any $\mathfrak{D}^2$, let $x \in \mathbf{A} \in \mathfrak{D}^2$ be an element of $CS_{LIB\text{-}MFT}$ iff there exists at least one $C_i \in \mathbf{R}^2 \in \mathfrak{D}^2$ such that x is an element of CS_{P1} according to some $P^1 \in \mathbf{P}^1 \in \mathfrak{D}^2$ and there exists no $P^{1\prime} \in \mathbf{P}^1 \in \mathfrak{D}^2$ for which $C_i(P^{1\prime})>C_i(P^1)$; otherwise, x is an element of $PS_{LIB\text{-}MFT}$.

The informal definition references 'the (an) imprecise decision principle,' but we could also let it reference 'the (an) $(n\text{-}1)^{th}$-order decision principle,' like so

Liberal-My Favourite Theory (informal). x is permissible iff x is permissible according to the (an) $(n\text{-}1)^{th}$-order decision principle that is assigned highest credence relative to at least one credence function in the representor; otherwise, x is impermissible.

With regards to the formal definition, $\mathfrak{D}^2$ should be changed to $\mathfrak{D}^n$ and reference to some imprecise decision principle P^1 should be changed to reference to some $(n\text{-}1)^{th}$-order decision principles $P^{(n-1)}$, like so

Liberal-My Favourite Theory (formal). For any $\mathfrak{D}^n$, let $x \in \mathbf{A} \in \mathfrak{D}^n$ be an element of $CS_{LIB\text{-}MFT}$ iff there exists at least one $C_i \in \mathbf{R}^n \in \mathfrak{D}^n$ such that x is an element of $CS_{P(n\text{-}1)}$ according to some $P^{(n-1)} \in \mathbf{P}^{(n-1)} \in \mathfrak{D}^n$ and there exists no $P'^{(n-1)} \in \mathbf{P}^{(n-1)} \in \mathfrak{D}^n$ for which $C_i(P'^{(n-1)})>C_i(P^{(n-1)})$; otherwise, x is an element of $PS_{LIB\text{-}MFT}$.

Recall next the second-order decision principle called

Maximality-My Favourite Alternative (informal). x is permissible iff there is no y such that y is strictly preferred to x; otherwise, x is impermissible. Furthermore, x is strictly preferred to y iff x is more likely to be permissible than y relative to every credence function in the representor; x is indifferent to y iff x is as likely to be permissible as y relative to every credence function; otherwise, preferences between x and y are indeterminate.

Maximality-My Favourite Alternative (formal). For any $\mathfrak{D}^2$, let $x \in \mathbf{A} \in \mathfrak{D}^2$ be an element of $CS_{MAX\text{-}MFA}$ iff there is no $y \in \mathbf{A} \in \mathfrak{D}^2$ such that y is strictly preferred to x; otherwise, x is an element of $PS_{MAX\text{-}MFA}$. Furthermore, x is strictly preferred to y iff relative to every $C_i \in \mathbf{R}^2 \in \mathfrak{D}^2$, x is more likely to be an element of CS_R than y; x is indifferent to y iff relative to every C_i, x is as likely to be an element of CS_R as y; otherwise, preferences between x and y are indeterminate.

The informal definition works as given. In the formal definition, $\mathfrak{D}^2$ should be changed to $\mathfrak{D}^n$ and $\mathbf{R}^2$ to $\mathbf{R}^n$, as follows

Maximality-My Favourite Alternative (formal). For any $\mathfrak{D}^n$, let $x \in \mathbf{A} \in \mathfrak{D}^n$ be an element of $CS_{MAX\text{-}MFA}$ iff there is no $y \in \mathbf{A} \in \mathfrak{D}^n$ such that y is strictly preferred to x; otherwise, x is an element of $PS_{MAX\text{-}MFA}$. Furthermore, x is strictly preferred to y iff relative to every $C_i \in \mathbf{R}^n \in \mathfrak{D}^n$, x is more likely to be permissible than y; x is indifferent to y iff relative to every C_i, x is as likely to be permissible as y; otherwise, preferences between x and y are indeterminate.

The other second-order decision principles that I've developed can be translated into n^{th}-order decision principles for any n>2 in similar ways.

Condition (v). To determine which alternatives are in fact permissible or impermissible (i.e., elements of CS_R or PS_R, respectively), rational uncertainty about n^{th}-order decision principles for any n should be handled by invoking an $(n+1)^{th}$-order decision principle.

More concretely, rational uncertainty about first-order decision principles, i.e., imprecise decision principles, should be handled by invoking a second-order decision principle; rational uncertainty about second-order decision principles should be handled by invoking a third-order decision principle; and so on.

That rational uncertainty about imprecise decision principles should be handled by means of a second-order decision principle is something that any Metanormativist should accept. That uncertainty about higher order decision principles should be dealt with using even higher order decision principles is essential to any Hierarchy Account. It's also essential to the Metanormativist who has given up on Unique Accounts. For starters, that rational uncertainty about second-order decision principles should be handled by means of a third-order decision principle is entailed by the failure of Unique Accounts. Since there isn't any unique second-order decision principle that decision-makers should conform to, and since decision-makers can be rationally uncertain about such principles, decision-makers need some other sort of principle to come to a decision: A third-order decision principle.

At this point, the Metanormativist might claim that there's a unique third-order decision principle, one that decision-makers should always conform to in order to account for uncertainty about second-order decision principles (see §3.1).

However, as we've just seen, third-order decision principles aren't significantly different from second-order decision principles: The third-order decision principles of which we're (or, at least, I'm) aware are reworked second-order decision principles. As a result, the arguments that we have used to argue that there's no unique second-order decision principle that decision-makers should always obey can be put to work to argue that there's no unique third-order decision principle that decision-makers should always obey (see §4.4). In brief: Most importantly, third-order decision principles are variations on imprecise decision principles, making it implausible that uncertainty about them shouldn't be accounted for in decision-making on the assumption that uncertainty about imprecise decision principles should. Furthermore, each of the third-order decision principles can be subjected to criticism, there's no explanation for why a single third-order decision principle should always be used, even in the face of uncertainty about it, and the arguments in favour of Metanormativism sketched in §3.1 work equally well to defend the claim that rational uncertainty about third-order decision principles should rationally be accounted for in decision-making. If the Metanormativist wants to claim that there's a third-order decision principle that doesn't have these problems, the burden of proof lies with her.

These same arguments show that there's no single fourth-order decision principle that should always be used, and that uncertainty about fourth-order decision principles should be accounted for using some other sort of decision principle, namely a fifth-order decision principle; and so on. So, Condition (v) should be accepted.

> *Condition (vi).* The decision-maker's credences towards the same decision principles as employed at different orders *cohere* with each other.[22]

Making this more precise requires spelling out the notion of 'coherence,' which I'll not attempt. An example should hopefully get the point across. Suppose that a decision-maker gives credence to Liberal-MFT as an n^{th}-order decision principles for any $n>1$. More specifically, suppose that she gives credences ranging from 0.5 to 0.8 to Liberal-MFT as a second-order decision principle. At the same time, she gives credence 0 to Liberal-MFT as a third-

[22] I'm here building on comments by MacAskill (2016b, 444) and Trammell (2021, 1188).

order decision principle, on every credence function in her representor; credences ranging from 0 to 1 to Liberal-MFT as a fourth-order decision principle; and so on. This decision-maker's credences towards Liberal-MFT as employed at every order *n* don't seem to *cohere* with each other. If the credences ranging from 0.5 to 0.8 to Liberal-MFT as a second-order decision principle are rationally assigned, then there must be reason for giving them. But it seems odd that this reason wouldn't also pertain to the principle as defined for higher orders, in which case higher order versions should receive the same credence (i.e., credences ranging from 0.5 to 0.8), or at least similar credence (e.g., perhaps it's allowed to give credences ranging from 0.45 to 0.82). Given the tight connection between rationality and coherence (see §§2.1-2.2), any rational decision-maker's credences towards the same decision principles as employed at different orders should cohere with each other. (A similar claim was implicit when I held that rational uncertainty about imprecise decision principles entails that one should be rationally uncertain about second-order decision principles, given that these are variations on imprecise decision principles; see §4.4).

So, any rational decision-maker who gives credence to second-order decision principles P_i^2 and P_j^2, and who is aware of the fact that each can be naturally reworked as an *n*th-order decision principle for any *n*>2, should give credences pertaining to P_i^n and P_j^n as *n*th-order decision principles for any *n*>2 that cohere with the credences given to P_i^2 and P_j^2. For present purposes, it's of particular importance that the decision-maker's credences are such that at no order *n*, P_i^n or P_j^n is given credence 0 or credence 1 relative to every credence function in the decision-maker's representor. Furthermore, at no order *n*, the decision-maker can withhold judgment about P_i^n or P_j^n. This is eminently reasonable–for the decision-maker's credences to be coherent much more is required (e.g., if the decision-maker gives credences ranging from 0.5 to 0.8 to P_i^2, she should give the same or similar credences to P_i^n for any *n*>2).

Finally, there's

> *Condition (vii).* At every order *n*, (a) P_i^n agrees with P_i^1 that only x is permissible, and all other available alternatives are impermissible; (b) P_j^n agrees with P_j^1 that only y is permissible, and all other available alternatives are impermissible.

If these seven conditions hold in a single case, infinite regress arises, such that at every order, the decision principles to which the decision-maker gives credences disagree about the permissibility of the alternatives. I'll now show this. By Condition (ii), the decision-maker faces a decision under cluelessness, such that imprecise decision principle P_i^1 says that only x is permissible, and all other available alternatives are impermissible, and imprecise decision principle P_j^1 says that only y is permissible, and all other available alternatives are impermissible. By Condition (iii), the only imprecise decision principles to which the decision-maker gives credence are P_i^1 and P_j^1. By Condition (i), the decision-maker should take her uncertainty about P_i^1 and P_j^1 into account to come to a rational decision. By Condition (iv), P_i^2 and P_j^2 are the only second-order decision principles to which the decision-maker gives credence. By Condition (v), rational uncertainty about P_i^2 and P_j^2 should be handled by invoking a third-order decision principle. By Condition (iv), the decision-maker is aware of the fact that P_i^2 and P_j^2 can be naturally reworked as third-order decision principles. By Condition (vi), the decision-makers credences towards P_i^2 and P_i^3 as well as towards P_j^2 and P_j^3 should cohere, which at the very least means that she should be uncertain about whether to conform to P_i^3 or P_j^3. By Condition (viii), P_i^3 agrees with P_i^1 that only x is permissible, and all other available alternatives are impermissible; P_j^3 agrees with P_j^1 that only y is permissible, and all other available alternatives are impermissible. The last four steps can theoretically be repeated *ad infinitum*, leading to the need for ever higher order decision principles. And so we have arrived at an infinite regress, such that at every order, the decision principles to which the decision-maker gives credence disagree about which alternatives are permissible. Specifically, the principles disagree about whether x or y is permissible.

Following MacAskill, Bykvist, and Ord (2020, 31), one might object that real-life decision-makers are cognitively limited and, therefore, cannot be uncertain about n^{th}-order decision principles for every $n>2$. If there's some upper limit, this would ward off infinite regress.

My reply is this: The conditions should indeed be able to apply to real-life decision-makers. I readily admit that real-life decision-makers can't ever consider every order individually: It's of course impossible to entertain second-order decision principles and rationally conclude that you're uncertain about them, move on to the third order, and go on like this until you've entertained all n orders for every $n>2$. However, this isn't necessary to attain infinite regress. In the case under consideration, it's important that the decision-maker is *aware* that the two second-order decision principles to which she gives credence can be employed at every order above (as stipulated by Condition (iv))). If so, it should as a result become irrational for her to be certain about any of these principles, or to withhold judgment about them, at any higher order (as stipulated by Condition (vi)). But there's no need for her to consider every order individually to end up in a state of uncertainty about decision principles at every order–we've not considered every order individually in the argument above either, and yet the infinite regress was attained. So, normal human cognitive limitations don't stand in the way of infinite regress.

§5.3 Permissibility in the Face of Regress

So, then, Hierarchy Accounts should allow for a case in which decision principles at all orders disagree about whether x or y is permissible. In this subsection, I'll argue that, in this case, it's left unspecified whether any alternative is permissible, i.e., an element of CS_R, due to which $CS_R=\{\}$. In short, the argument is this: In a case where Conditions (i)-(vii) are met, saying that x, y, or x and y are permissible requires appeal to some sort of decision principle, and the Metanormativist isn't in the position to identify any principle fit do the task.

For starters, suppose that we try to appeal to P_i^n for some *n* to claim that only x is permissible (or, alternatively, to P_j^n for some *n* to claim that only y is permissible). Since the decision-maker is uncertain whether she should abide by P_i^n or P_j^n, abiding by the former and not the latter would be arbitrary. The Metanormativist should say that her uncertainty should be accounted for in decision-making, using some (n+1)th-order decision principle. But, she should then use either $P_i^{(n+1)}$ or $P_j^{(n+1)}$, about which she's uncertain. If the Metanormativist wishes to insist that it *can* be rational for the decision-maker to conform to P_i^n if she gives it much greater credence than P_j^n, let's stipulate that she gives the principles (roughly) equal credence (this shouldn't conflict with Conditions (i)-(viii), at least not for all ways in which they can be fleshed out).

Next, the Metanormativist might want to appeal to some decision principle that isn't part of the infinite hierarchy, but that's somehow external to it, ranging over all the principles are part of the hierarchy (see §3.1). But, whatever the content of this principle, we can suppose that the decision-maker is uncertain about it, in which case the Metanormativist is reasonably committed to the claim that uncertainty about this principle should be accounted for in decision-making. We can suppose that while some 'external' decision principles to which the decision-maker gives credence declare only x permissible, others declare only y permissible.

Finally, one might hold that if we cannot privilege one alternative over the other, x and y should both be declared permissible. But, how might we justify this position? Clearly, we need one or other decision principle according to which this is so. But, of course, the decision-maker might rationally be uncertain about this principle. Since her uncertainty should be taken into account in decision-making, we can end up with yet another infinite regress with some principles declaring both x and y, permissible, some only x permissible, and others only y permissible.

Perhaps the Metanormativist will try to argue that x, y, or x and y are permissible via some other route. But the point is clear: She cannot escape from the conclusion that $CS_R=\{\}$.

One might object to the above argument that I'm setting the bar too high. As an example, consider again the appeal to P_i^n for some *n* to claim that only x is permissible. One might hold that even if the decision-maker isn't certain about P_i^n, it's rational for her to rely on this principle to conclude that x is permissible. Otherwise, I'd set an impossible standard for the Metanormativist. Riedener (2021, 16–17) hints at this objection (the objection actually raised by Riedener is more subtle, but the subtleties don't apply to the present context, so I'll set these aside).

I'm not convinced. To make sure that we're all on the same page, I'm not claiming that there's *no* reason for the decision-maker to choose x, or that choosing x by appealing to P_i^n would be arbitrary through and through. What I'm claiming is that x isn't all-rationally-relevant-things-considered rationally permissible. The same goes for y. This is considerably weaker. Furthermore, with regards to the claim that "I'm setting the bar too high" or that I'm setting "am impossible standard for the Metanormativist," I don't think I am. What I'm doing is, basically, using Metanormativism against itself (by, roughly, extending the claim that uncertainty about imprecise decision principles should be accounted for in decision-making to the claim that uncertainty about decision principles in general should be accounted for in decision-making). That's fair game.

So, in sum, in the case under consideration, CS_R is empty.

§5.4 Decisiveness

Some Metanormativists might not find this troubling, and will acknowledge that there are cases of infinite regress in which no alternative is declared permissible; see, e.g., MacAskill (2014, 219). Nevertheless, I'll argue that, if we make one final addition to the case under consideration, it's fatal to any remotely plausible Hierarchy Account.

Given that in the case we've been discussing, $CS_R=\{\}$, the Hierarchy Account must, at least in this particular case, violate

> *Decisiveness (informal).* In any decision-situation, at least one available alternative is permissible.
>
> *Decisiveness (formal).* In any decision-situation $\mathfrak{D}$, at least one $x \in \mathbf{A} \in \mathfrak{D}$ is such that $x \in CS_R$, i.e., $CS_R \neq \{\}$.

Although the Hierarchy Account might still uphold

> *Weak Decisiveness (informal).* In any decision-situation, at least one available alternative is not impermissible.
>
> *Weak Decisiveness (formal).* In any decision-situation $\mathfrak{D}$, at least one $x \in \mathbf{A} \in \mathfrak{D}$ is such that $x \notin PS_R$, i.e., $PS_R \neq \mathbf{A}$.

Hence, the crux is whether it's acceptable to the Metanormativist that Decisiveness cannot be satisfied in this case, and whether she can make do with Weak Decisiveness instead.

Some decision principles meet both conditions (e.g., Liberal, Maximality). Others satisfy only Weak Decisiveness. Moderate is such a principle: Because it will sometimes declare all alternatives *indeterminately* permissible or impermissible, in some decision-situations no alternative will be declared permissible (see, e.g., Nina's case). Some principles satisfy neither (Joyce 2010, 314 considers an imprecise decision principle that violates both–'Conservative'–, but rejects it on pain of incoherence, precisely because it effectively violates both conditions).

So far, I've left the nature of imprecise decision principles P_i^1 and P_j^1 and n^{th}-order decision principles P_i^n and P_j^n for any n>1 open. Let's suppose that all these principles meet Decisiveness. In case the decision-maker also gives credence to decision principles that are external to the hierarchy (see the previous subsection), suppose that all these meet Decisiveness as well.

(To be sure, this doesn't require the decision-maker to be certain that 'non-decisive' decision principles are false; we can suppose that she's just not aware of any principle that violates Decisiveness). So, this decision-maker can be rationally certain that a decision principle that meets Decisiveness is true and, so, certain that Decisiveness is true. Nevertheless, if what she should rationally do depends on her uncertainty about decision principles–in her case, a state of rational uncertainty that entails that Decisiveness must be true–, then Decisiveness must be false. That is, the Hierarchy Account comprised exclusively of 'decisive' principles itself fails to meet Decisiveness. This is unacceptable, for two reasons. So, there's no adequate Hierarchy Account.

§5.4.1 *Argument 1 for the Unacceptability of Decisiveness Violation: Contradictory Credences*

The first reason that it's unacceptable that a Hierarchy Account comprised exclusively of 'decisive' principles itself fails to meet Decisiveness is this. Suppose that the decision-maker were to embrace the Hierarchy Account. That is, she comes to give considerable credence to the proposition that the whole hierarchy of decision principles to which she gives credence comprises the account of choice on which she must act. If this were rationally allowed, she'd be rationally allowed to give credence (i.e., positive credence) to the proposition that Decisiveness is violated, at least in this case. However, at the same time, she can be rationally certain that Decisiveness is true and, therefore, that this condition cannot possibly be violated by the true account of rational choice. So, allowing her to give credence to the Hierarchy Account means allowing her credences to be contradictory. This cannot be rationally acceptable. Hence, she cannot rationally embrace the Hierarchy Account. But if it were true, she'd be able to rationally do so. Therefore, it's false.

Here's a reply that might seem persuasive at first sight: In some circumstances, rationality denies that you can give credence to true propositions. For example, if all the available evidence tells you to believe that you're reading a philosophy thesis, even though you're really a brain in a vat, it's rationally prohibited (we can suppose, for argument's sake) to give credence to the true proposition that you're a brain in a vat. This is a similar occasion. The decision-maker simply cannot give credence to the Hierarchy Account, but this doesn't mean it's false.

The Metanormativist, however, cannot avail herself of this reply. There's a significant difference between cases where rationality denies that can you give credence to true propositions (as in the brain-in-a-vat scenario), and the case of the Hierarchy Account.

In cases of the former type, if we were to reject all *misleading* available evidence, i.e., available evidence against true propositions, we'd end up in a state where it might be rational to give credence to true propositions. For instance, if Dr. Evil would stop letting it seem to you that you're reading a philosophy thesis, you'd be rationally allowed to give credence to the true proposition that you're a brain in a vat.

With regards to the Hierarchy Account, the relevant misleading evidence pertains to the false imprecise decision principles that (in part) make up the Hierarchy Account. Any Hierarchy Account is necessarily build up from infinitely many false decision principles (given that, at every order, the decision principles are assumed to be mutually exclusive; see §3.2). It's the misleading evidence that (in part) lets the Hierarchy Account come into existence, so to speak. Hence, if the decision-maker were to reject all misleading evidence, the Hierarchy Account itself would fall apart–she'd not be allowed to give the account credence under these circumstances either, given that she rejects some of its elements.

So, there's a difference between the case of the Hierarchy Account and cases in which rationality denies that you can give credence to true propositions. Therefore, the proponent of the Hierarchy Account cannot draw on the latter kind of case to argue that it shouldn't be possible for the decision-maker to rationally give credence to the Hierarchy Account, even if it's true. Thus, if the Hierarchy Account were true, the decision-maker (who gives credence to infinitely many decision principles that jointly make up the Hierarchy Account), should be able to rationally give it credence. That's impossible, and so the Hierarchy Account cannot be true.

§5.4.2 Argument 2 for the Unacceptability of Decisiveness Violation: Condition Dependence

The second reason that it's unacceptable for a Hierarchy Account comprised exclusively of 'decisive' principles to violate Decisiveness is as follows. It may be acceptable for Metanormativists to take *some* conditions for granted, i.e., to hold that decision-makers must necessarily conform to some conditions in order to be rational, no matter their attitudes. An example might be the law of non-contradiction. Even if decision-makers are unaware of the law of non-contradiction or deny it, the Metanormativist might reasonably claim that it can be rationally imposed on decision-makers. However, this of course doesn't hold for every true condition, else uncertainty about true imprecise decision principles needn't be accounted for in decision-making. The Metanormativist might suggest that the negation of Decisiveness is like the law of non-contradiction: It can be imposed on decision-makers, even if they deny it (i.e., hold that Decisiveness is true).

But, the negation of Decisiveness makes for an incredibly poor candidate for the type of condition that can, according to the Metanormativist, be imposed on decision-makers, no matter their attitudes. This is because any remotely plausible Hierarchy Account should respect

> *Condition Dependence*. Whether a Hierarchy Account should entail a condition that is met by some, but not all, decision principles, depends on whether this condition is met by the decision principles to which the decision-maker gives credence, i.e., the decision principles that jointly form the Hierarchy Account. In particular, if all (none) of the decision principles to which the decision-maker gives credence meet a condition, this condition should (not) be entailed by the Hierarchy Account.

The law of non-contradiction doesn't conflict with any imprecise decision principle, so Condition Dependence says that it should be entailed by any Hierarchy Account. So far so good. But to see just how plausible Condition Dependence is, consider the following intuition pump to get you in the right mindset. As noted in §2.4, Elga (2010) has argued that there's no plausible imprecise decision principle. Elga's argument centres on a decision-maker who values all and only monetary gains. Elga asks us to suppose that she is offered the following series of bets regarding any proposition H, in short succession (as a result of which, her evidence regarding H and her interpretation of this evidence remains unchanged):

> Bet A: If H is true, you lose \$10. Otherwise you win \$15.
> Bet B: If H is true, you win \$15. Otherwise you lose \$10. (Elga 2010, 4)

If the decision-maker accepts both bets, she will gain \$5. Elga points out that this does not imply that the decision-maker is rationally required to do so; it may be that she is so confident that H is true, that it's rational for her to only accept Bet B. Elga claims that the assured gain of \$5 in case she accepts both bets, however, does entail that she is rationally required to accept *at least one* bet. Elga notes that decision principles that use precise credences, such as MEV, are able to accommodate this requirement (4–5). But, he argues, some imprecise decision principles permit decision-makers to reject both bets (4–6).

This includes Liberal. Consider the decision about whether to accept or reject Bet A. Assume that the decision-maker's representor contains some credence function relative to which rejecting Bet A has greater expected value than accepting it (this will be so, e.g., if H receives credence 0.9). If so, rejecting Bet A maximizes expected value relative to at least one credence function in the representor. Hence rejecting Bet A is permissible according to Liberal. Similar reasoning, involving, e.g., a credence function where H receives credence 0.1, shows that she is permitted to reject Bet B as well. So, Liberal allows her to reject both bets, thereby foregoing a certain gain of $5. Just as Liberal, Maximality fails Elga's (2010) test (as its proponents Herlitz and Mogensen are well aware; see Herlitz 2019, 14–15; Mogensen 2021, 149–150), although I won't show it here.

Moderate fares better on Elga's task (Rinard 2015, 6). Consider once more the decision about whether to accept or reject Bet A. Suppose that the decision-maker's representor contains some credence function relative to which *rejecting* Bet A has greater expected value than accepting it (e.g., if H receives credence 0.9) and, furthermore, some credence functions relative to which *accepting* Bet A has greater expected value than rejecting it (e.g., if H receives credence 0.1). Moderate says that its *indeterminately permissible* to reject Bet A as well as indeterminately permissible to accept it. Similar reasoning, involving similar credence functions (e.g. with H receiving credence 0.9 and credence 0.1), shows that the decision-maker is indeterminately permitted to accept and to reject Bet B. In sum, the decision-maker isn't permitted to reject both bets, although it is indeterminately permissible for her to do so.

Now, why is all this relevant? Given that what decision-makers ought to do depends on their credences given to imprecise decision principles (as Metanormativism claims), it would be odd if it would be necessarily irrational for decision-makers to reject both bets. Consider a decision-maker who is certain that either Maximality or Liberal is correct. According to any remotely plausible Metanormativist account, she ought to be permitted to reject both bets, simply because all imprecise decision principles to which she gives credence permit her to do so (all second-order decision principles presented in §4.3 would tell her that, indeed, she's permitted to reject both bets). However, if a decision-maker were to give credence only to Moderate and other decision principles that would say that decision-makers shouldn't be permitted to reject both bets, then any plausible Metanormativist account would agree. Hence, whether a Hierarchy Account should entail the condition that decision-makers shouldn't be permitted to reject both bets, depends on whether this condition is met by the decision principles to which the decision-maker gives credence. This is precisely the verdict delivered by Condition Dependence. So, it's an attractive requirement.

With regards to Decisiveness: As noted at the start of this subsection, Liberal and Maximality satisfy this condition, but it's violated by Moderate (which does meet Weak Decisiveness). If decision-makers give credence only to 'decisive' decision principles, Condition Dependence entails that Hierarchy Accounts should entail Decisiveness. So, Condition Dependence says that, in the case under consideration, the Hierarchy Account should respect Decisiveness. (Of course, if decision-makers give credence only to 'weakly decisive' decision principles, such as Moderate, Condition Dependence demands that Decisiveness is violated. But this is what we should expect from any remotely plausible Hierarchy Account). If the Metanormativist wants to claim that Decisiveness is exempt from Condition Dependence, i.e., if Condition Dependence *does* apply to the condition that decision-makers shouldn't be permitted to reject both Elga's bets, but *doesn't* apply to Decisiveness, she needs to provide us with some plausible explanation. To claim that Decisiveness should be denied purely in response to the infinite regress argument would be *ad hoc*. I doubt there's such an explanation, but the burden of proof rests with the Metanormativist.

For this and the previous reason, the Metanormativist cannot reasonably accept that, in the case under consideration, Decisiveness is violated by the Hierarchy Account. This renders such accounts incoherent. One the one hand, Hierarchy Accounts cannot accept the violation of Decisiveness in this case. On the other, we've seen that, in this case, Hierarchy Accounts must accept the violation of Decisiveness. And so the infinite regress is vicious: It shows that there's no plausible Hierarchy Account. To wrap up, I'll respond to potential Metanormativist counterarguments.

§5.5 Prolepsis

I'll consider four responses to the infinite regress argument on behalf of the Metanormativist, in no particular order. None succeeds.

§5.5.1 Objection 1 to the Infinite Regress Argument: Odd Epistemic Situations

One might–following MacAskill (2014, 219)–hold that it's to be expected that in some cases of infinite regress, it will be left unspecified whether any alternative is permissible. The decision-maker is just "in an odd epistemic situation where there's no way of aggregating her uncertainty such that" any alternative can be reasonably be called permissible (idem).

My reply is this: The decision-maker in the case under consideration is surely in an odd epistemic situation, but not one where it can reasonably be expected that no alternative can be called permissible. Her epistemic situation is (as we've seen) such that she can be certain that at least one alternative is permissible.

§5.5.2 Objection 2 to the Infinite Regress Argument: Appeal to Rinard and Moderate

As noted in §5.4, Moderate doesn't satisfy Decisiveness–because it will sometimes declare all available alternatives indeterminately permissible–, although it does satisfy Weak Decisiveness. Rinard (2015)–who, recall, developed and defended Moderate–is aware of this.[23] Rinard doesn't find it troubling, and argues against those who might (13–14). The Metanormativist could try to borrow Rinard's line of reasoning, and use it to defend the Hierarchy Account's violation of Decisiveness.

Rinard's argument runs roughly as follows. The appeal of Decisiveness derives from its seeming necessity for decision principles to satisfy 'ought implies can.' In cases where no available alternative is permissible, it may seem as if decision principles require decision-makers not to choose *any* alternative. This is impossible, and requirements to do something impossible go against 'ought implies can.' However, Rinard argues, to satisfy 'ought implies can,' it suffices that decision principles meet Weak Decisiveness. As long as not all available alternatives are rationally impermissible (as Weak Decisiveness requires), some alternatives will be permissible or indeterminately permissible. As a result, decision principles will not tell the decision-maker not to choose any alternative. Hence, they will not require the decision-maker to do something impossible, and 'ought implies can' isn't violated. The proponent of the Hierarchy Account might claim that this argument shows that the Hierarchy Account can violate Decisiveness, too.

My reply is two-fold: Firstly, I'm not denying that Decisiveness is necessary to satisfy 'ought implies can.' So, Rinard's argument isn't of much help to the Metanormativist. Secondly, the proponent of Moderate can provide us with a plausible explanation of why this imprecise decision principle fails to meet Decisiveness. If alternatives can be indeterminately permissi-

[23] Although Rinard doesn't use the labels 'Decisiveness' or 'Weak Decisiveness.'

ble, and, moreover, there are decision-situations where Moderate says that all available alternatives are indeterminately permissible or impermissible, then there must be cases where no alternative is permissible, i.e., Decisiveness must be false (this hangs on the acceptability of 'indeterminate permissibility,' but as noted in §2.5, delving into the justification of this notion is beyond the scope of this thesis). Hierarchy Accounts don't have this luxury, not in the case we've been considering: Given that all decision principles that comprise the Account are 'decisive,' it's fair to stipulate that none acknowledge the possibility of 'indeterminate permissibility.' As noted in §5.4, to claim that Decisiveness should be denied by Hierarchy Accounts purely in response to the infinite regress argument would be *ad hoc*. I doubt that there's a suitable explanation for the failure of Hierarchy Accounts to respect this condition even if all decision principles that comprise the account satisfy it. If the Metanormativist wants to claim that there is such an explanation, it's up to her to find it.

§5.5.3 Objection 3 to the Infinite Regress Argument: Jackson Cases

The Metanormativist might argue that there's an analogy between the infinite regress argument (on the one hand) and so-called *Jackson cases* (on the other), which shows that it's not problematic for the Hierarchy Account to reject Decisiveness in the case under consideration. To see what Jackson cases are, let's consider one of the first, due to (who else) Jackson (1991):

> *The Drug Example, Mark 1*. Jill is a physician who has to decide on the correct treatment for her patient, John, who has a minor but not trivial skin complaint. She has three drugs to choose from: drug A, drug B, and drug C. Careful consideration of the literature has led her to the following opinions. Drug A is very likely to relieve the condition but will [certainly] not completely cure it. One of drugs B and C will completely cure the skin condition; the other though will kill the patient, and there is no way that she can tell which of the two is the perfect cure and which the killer drug. What should Jill do? (Jackson 1991, 462–463).

One the one hand, Jill is certain that she shouldn't prescribe drug A, for it's certainly not the drug that will cure John completely. On the other, she should nevertheless clearly prescribe it, since B and C are off the table. There's no real contradiction here; we can make sense of both claims.

The Metanormativist might say that something similar is going on in the case we've been considering. On the one hand, the decision-maker is certain that Decisiveness is met, for it's satisfied by all decision principles to which she gives credence. On the other, Decisiveness is clearly violated. Even so, there's no real contradiction.

My reply is this: On first sight, this seems persuasive. However, once we dig a just little deeper and see what Jackson cases actually show us, the analogy with the case of the Hierarchy Account breaks down.[24] According to *objective* consequentialism, *x* is permissible iff *x in fact* maximizes the good. On this view, Jill should either prescribe drug B or C; whichever is the perfect cure. Jill shouldn't prescribe drug A, since it will in fact bring about a worse outcome than the perfect cure. It's in the sense of objective consequentialism (or some other 'objective' principle) that Jill is certain that she shouldn't prescribe drug A. Next, according to *subjective* consequentialism (on the assumption that precise credences are assigned to outcomes), *x* is permissible iff *x* maximizes the *expected* good. On this view, Jill should prescribe drug A. It's in the sense of subjective consequentialism (or some other 'subjective' principle) that Jill clearly should prescribe drug A (see 465–466).

This sort of move is unavailable to the Metanormativist. While she can claim that 'in the sense of some decision principles,' Decisiveness is met, she can't say that 'in the sense of some other decision principle(s),' Decisiveness isn't met. After all, the decision-maker in the case under consideration is certain that every decision principle meets this condition. So, there isn't a proper analogy between Jackson cases and the case of the Hierarchy Account from which the Metanormativist might draw to reject the infinite regress argument.

§5.5.4 Objection 4 to the Infinite Regress Argument: The Silence of Rationality

The Metanormativist might object that while there *is* an account of rational decision-making that takes rational uncertainty about decision principles into account, it *isn't* always applicable. That is, sometimes rationality is *silent.* In cases such as the one that featured in the infinite regress argument, there's no fact of the matter about whether alternatives are rationally permissible. So, the Hierarchy Account isn't rendered incoherent.

---

[24] Interestingly, Jackson cases are also used to show that certain decision principles aren't sufficiently *action guiding* (see, e.g., Jackson 1991, 466–467; MacAskill and Ord 2020, 330–332).

My reply is two-fold: First, claiming that rationality is sometimes silent should be a last resort. If there's an account of rational decision-making available on which rationality isn't silent, we should prefer it. And we've seen such accounts: Imprecise decision principles of the sort discussed in §§2.4-2.5 specify whether alternatives are rationally permissible or impermissible in *every* decision-situation. Second, the claim that the Hierarchy Account isn't always applicable needs to be properly motivated. Otherwise, this claim would be *ad hoc*. The burden of proof rests, per usual, with the Metanormativist.

§5.6 Conclusions about Hierarchy Accounts

In sum, then, the infinite regress argument shows that there's no adequate Hierarchy Account. Since there's also no adequate Unique Account, the upshot is that

> *Metanormativism about Imprecise Decision Principles.* Rational uncertainty about imprecise decision principles should rationally be accounted for in decision-making.

is false. To wrap up, I'll tease out some implications of this conclusion for effective altruism.

§6. Implications for Effective Altruism

I've spend the past several sections arguing against Metanormativism. However, what got us started was effective altruism. To recap, quickly: Many effective altruists endorse MEV for the sake of rational action guidance, i.e., to help them identify which actions they should perform, rationally speaking, in order to do the most good. However, given that effective altruists are frequently confronted with decisions under cluelessness, they require some decision principle to replace it. Decision theorists have thought long and hard about rational decision-making in cases where imprecise credences are assigned to outcomes conditional on choosing the alternatives that may bring them about, and so it seems that one or other imprecise decision principle should take MEV's place. Yet it can be rational to be uncertain about such principles. Now we've seen that this type of uncertainty isn't relevant for rational decision-making.

What does this mean, practically speaking, as far as effective altruism goes? In this section, I'll set out my current thoughts on the implications of the falsehood of Metanormativism for effective altruism. All of it is preliminary, most of it is tentative, and little will satisfy the typical effective altruist.

Now, it seems that the falsehood of Metanormativism doesn't entail that effective altruists shouldn't factor in *any* uncertainty into their decision-making (cf. MacAskill 2014, 218). In particular, it still seems reasonable to take uncertainty about outcomes into account. Given that this uncertainty is best modelled using imprecise credences, this would mean that cluelessness is relevant for decision-making. When taking uncertainty about outcomes into account, it's essential that these outcomes don't vary with respect to imprecise decision principles. That is, for every decision-situation $\mathfrak{D}$, we should assume that every $o_i \in O \in \mathfrak{D}$, is either (i) such that the same imprecise decision principles are true (e.g., that Liberal is true in every outcome) and false (e.g., that Moderate and Maximality are false in every outcome), or (ii) such that imprecise decision principles are left out of the outcome (description) entirely (e.g., no reference is made to Liberal, Moderate, or Maximality in any outcome). If neither clause (i), nor clause (ii), is respected, uncertainty about imprecise decision principles may be taken into consideration in decision-making. It seems fair to presume that the failure of Metanormativism entails that uncertainty about second- and higher order decision principles is irrelevant for decision-making as well. If so, outcomes shouldn't vary with respect to such decision principles either.

How, then, should effective altruists take uncertainty about outcomes–that don't vary with respect to imprecise, second- or higher order decision principles–into account in their decision-making? Well, we've seen that decision theorists have proposed various imprecise decision principles that can be used for this very purpose. So, some imprecise decision principle seems a suitable candidate. Somewhat vaguely, we can say that it should be an *adequate*, or *true*, imprecise decision principle (it surely shouldn't be an imprecise decision principle that is inadequate, or false). But what makes for an adequate imprecise decision principle? Given that uncertainty about imprecise decision principles is irrelevant for rational decision-making, we cannot reasonably say that an imprecise decision principle is adequate because we assign it high credence–if that were so, then MFT wouldn't have been so implausible. Hence, it seems that the adequacy of imprecise decision principles doesn't depend on our *attitudes* towards them. If I'm on the right track, then the adequacy of imprecise decision principles doesn't appear to be a subjective affair. It's not so that an imprecise decision principle is adequate because it's reasonable to adopt given the available evidence. Rather, there will be some 'objectively adequate' imprecise decision principle, one that should be conformed to no matter our perspective on it.

But, of course, this claim doesn't undo or resolve any uncertainty about which imprecise decision principle to follow. It's utterly unclear which imprecise decision principle deserves to be called objectively adequate. So one thing that effective altruists could aim for is the reduction of uncertainty about imprecise decision principles. In case they rationally reach full certainty, they will have presumably identified the objectively adequate imprecise decision principle. Full certainty is, however, unattainable. Perhaps it's possible to, say, give credence 0.75 to some imprecise decision principle (on every credence function in one's representor). But this isn't enough to warrant labelling it the objectively adequate imprecise decision principle. Acting in accordance with such an imprecise decision principle looks suspiciously like acting in compliance with MFT, which, as we've seen, is something effective altruists shouldn't rationally do. This suggests that trying to reduce uncertainty about imprecise decision principles isn't a particularly promising path for the effective altruist. This is a strange conclusion. On the one hand, I'm saying that effective altruists should rationally conform to an objectively adequate imprecise decision principle, and that it's unclear to them what this principle is. On the other, I'm saying that it isn't of much use for them to try to figure out what this principle is, either. This is puzzling, but I'm not sure how to reasonably work around it, or whether that's even possible.

One might object that it cannot be rationally expected of effective altruists to abide by an imprecise decision principle in the event that they don't have a clue what it is. It may seem a violation of 'ought implies can.' But I'm not, strictly speaking, setting an impossible task: Any imprecise decision principle will say that certain available alternative are permissible or impermissible, and it's *always* possible for effective altruists to choose alternatives that are available to them. So, there's no violation of 'ought implies can,' and it's possible for effective altruists to follow the guidelines of an objectively adequate imprecise decision principle, even if they have no way of knowing what it and its prescriptions might be.

This isn't a very satisfying reply, of course, and one might press on by pointing out that an objectively adequate imprecise decision principle, the nature of which is a mystery, fails to be action guiding. One might hold that I'm effectively leaving effective altruists in the dark while facing decisions under cluelessness.

Given the importance of action guidance to effective altruism, this is indeed a serious concern. None of the claims I've made have been especially useful with regards to action guidance, and it's admittedly difficult to give effective altruists *any* practically helpful advice on the basis of my conclusions. In fact, I'd go so far as to say that the failure of Metanormativism entails that effective altruists are often *unavoidably clueless* in their decision-making. So, yes, there's a tension between the falsehood of Metanormativism, one the one hand, and the need for action guidance, on the other. But does that mean that it's reasonable to reject my conclusions? That would be so if any remotely plausible decision principle would have to be action guiding. Some philosophers have held such a view (e.g., Jackson 1991, 466–467), but not all. For example, some philosophers endorse objective consequentialism, i.e., the view that x is permissible iff x *in fact* maximizes the good (in spite of Jackson cases; see §5.2.2). I'm inclined to side with those who deny the necessity of action guidance.

One might stick to one's guns and maintain that effective altruists need action guidance. It's essential to effective altruism that there's some way of figuring out how to do the most good, and this method cannot be unknowable.

Here's the most accommodating reply that I've been able to think of:[25] For the sake of action guidance, effective altruists who are uncertain about imprecise decision principles (and who don't want to ignore them altogether), will presumably need to rely on some sort of Metanormativist account of choice, be it a Unique or Hierarchy Account (perhaps adjusted or extended in one way or another, as I've noted is possible in §3.1). Perhaps this can be done in an appropriate manner *even if Metanormativism is false and all Metanormativists accounts of choice are inadequate.* I'll explain.

Let's suppose, purely for illustrative purposes, that Liberal is the objectively adequate imprecise decision principle. Notice that Liberal *doesn't* tell us to: (i) establish which alternatives are available; (ii) establish the expected value of each available alternative relative to every credence function in the representor; (iii) choose the (an) alternative that maximizes expected value relative to at least one credence function in the representor. Instead of making claims (i)-(iii), or something resembling them, Liberal gives a necessary and sufficient condition for the permissibility (impermissibility) of alternatives (i.e., x is permissible iff x maximizes expected value relative to at least one credence function in the representor; otherwise, x is impermissible), and *that's it.* Thus, Liberal makes no claims about how we are to identify permissible and impermissible alternatives; that is, it doesn't commit us to any particular method of deliberation.

Perhaps some sort of Metanormativist account–even if it's not a proper theory of rational choice–can serve this purpose. Consider, again purely for illustrative purposes, the second-order decision principle that I've labelled Liberal-MFT. Just as Liberal, it only gives a necessary and sufficient condition for the permissibility (impermissibility) of alternatives (i.e., x is permissible iff x is permissible according to the (an) imprecise decision principle that is assigned highest credence relative to at least one credence function in the representor; otherwise, x is impermissible). Nevertheless, effective altruists could try to consciously act in accordance with it by: (i) establishing which alternatives are available; (ii) establishing the expected value of each available alternative relative to every credence function in the representor; (iii) establishing the verdicts about the permissibility and impermissibility of the alternatives according to the imprecise decision principles to which they give credence; (iv) choose the (an) alternative that is permissible according to the (an) imprecise decision principle that is assigned highest credence relative to at least one credence function in the representor. It

[25] I'm drawing on comments by Herlitz (2019, 11).

might be that following this recipe for decision-making would let effective altruists identify more alternatives that Liberal deems permissible than they would otherwise identify. If this is correct, then there's still some room for taking uncertainty about imprecise decision principles into account in decision-making, although not in the manner originally conceived by the Metanormativist.

Hence, effective altruists could have the best of both worlds: They can rightfully reject Metanormativism and acknowledge that Metanormativist accounts are inadequate as accounts of rational choice, without having to give up on action guidance, since they're permitted to take their uncertainty about imprecise decision principles into account in decision-making, using a Metanormativist account as a recipe for decision-making.

However, it's an empirical matter whether using any Metanormativist account of rational choice would indeed let effective altruists identify (more) alternatives that the objectively adequate imprecise decision principle deems permissible (than otherwise). Perhaps not. More worryingly, establishing whether the use of a Metanormativist account is successful requires some sort of standard. Specifically, we'll need to know which alternatives are in fact permissible according to the objectively adequate imprecise decision principle. Since we don't know what this is, we cannot properly evaluate any recipe for decision-making! So, I'm not convinced, and I take it that the pessimistic conclusion that effective altruists are unavoidably clueless may be inescapable. Nevertheless, to avoid this result, some approach like the one I've sketched here may (as far as I'm aware) be the effective altruist's best bet.

§7. Conclusion

Effective altruists often face decisions under cluelessness. As a result, their favoured decision principle, MEV, is often inapplicable. It stands to reason that an imprecise decision principle should take its place. However, it's possible to be rationally uncertain about imprecise decision principles, and, in line with much recent literature, one might endorse Metanormativism, i.e., the view that rational uncertainty about imprecise decision principles should rationally be accounted for in decision-making. In this thesis, I've argued against Metanormativism. The Metanormativist account of choice might take one of two general forms. First, it could be a Unique Account, i.e., a single second-order decision principle. Second, it can take the shape of a Hierarchy Account, i.e., an infinite hierarchy of second- and higher order decision principles to which the decision-maker gives credence. I've shown that neither a plausible Unique Account exists, nor a plausible Hierarchy Account. Hence, there's no plausible Metanormativist account of choice, which means that Metanormativism is false. More speculatively, I've suggested that the falsehood of Metanormativism doesn't show that uncertainty about outcomes should be ignored in decision-making, and that to factor in such uncertainty (when best modelled using imprecise credences), effective altruists should conform to an 'objectively adequate' imprecise decision principle–even if it's utterly unclear what this principle or its prescriptions might be. Consequently, the failure of Metanormativism entails that effective altruists are often unavoidably clueless in their decision-making. In order to respond to the worry that this will leave effective altruists without action guidance, I've suggested (though not accepted) that Metanormativist accounts of choice–even though they aren't proper theories of rational choice–could nevertheless be used as recipes for decision-making.

To reach these conclusions, I've, among other things, developed a–somewhat rough–classification and overview of the various forms that Metanormativists accounts of rational choice might take (Unique Accounts, Hierarchy Accounts, as well as extensions and variations thereof), developed a formal framework with which Metanormativist accounts can be defined that allow for imprecise credences (where before the literature generally assumed that decision-makers assign only precise credences); shown that Metanormativism may be paradoxical (in fn. 15); proposed nine second- and higher decision principles that can handle precise as well as imprecise credences given to decision principles; pointed the reader to the literature on social choice theory for even more such principles; developed a novel infinite regress argument; and, as an aside, showed that Mogensen's (2021) argument for his 'permissive' conclusion pertaining to effective altruist decision-making in the face of cluelessness is undermined because it's an instance of reasoning by means of MFT.

This thesis leaves various issues for future research. The following strike me as most important. Most obviously, it remains to be seen what the exact implications of the falsehood of Metanormativism are for effective altruism. Since my remarks on this matter were preliminary (and tentative), this will need to be investigated further. With regards to the various ways in which Unique and Hierarchy Accounts can be extended and adjusted, my claim that the arguments that I've developed apply to any extended or adjusted Metanormativist account deserves greater scrutiny. Finally, the type of Metanormativism that I've considered is a very restricted one: Rational uncertainty about imprecise decision principles is but one sort of normative uncertainty. The name 'Metanormativism' is more usually reserved for the view that normative uncertainty in general, or decision-theoretic, moral, or prudential uncertainty in particular, should be taken into consideration in decision-making. The failure of Metanormativism as applied to uncertainty about imprecise decision principles casts doubt on Metanormativism as a general position. But we should avoid hastily concluding that *no* type of normative uncertainty is relevant to decision-making. It's interesting to investigate the extent to which the arguments from this thesis carry over to other forms of Metanormativism.

Bibliography

Askell, Amanda. 2019. 'Evidence Neutrality and the Moral Value of Information'. In *Effective Altruism*, by Amanda Askell, 37–52. Oxford University Press. https://doi.org/10.1093/oso/9780198841364.003.0003.

Bales, Adam. 2018. 'Indeterminate Permissibility and Choiceworthy Options'. *Philosophical Studies* 175 (7): 1693–1702. https://doi.org/10.1007/s11098-017-0930-5.

Berg, Amy. 2018. 'Effective Altruism: How Big Should the Tent Be?' *Public Affairs Quarterly* 32 (4): 269–87. https://doi.org/10.2307/26910000.

Berkey, Brian. 2021. 'The Philosophical Core of Effective Altruism'. *Journal of Social Philosophy* 52 (1): 92–113. https://doi.org/10.1111/josp.12347.

Bradley, Richard C. 2017. *Decision Theory with a Human Face*. Cambridge: Cambridge University press.

Bradley, Seamus, and Katie Steele. 2014. 'Should Subjective Probabilities Be Sharp?' *Episteme* 11 (3): 277–89. https://doi.org/10.1017/epi.2014.8.

Bruers, Stijn. 2019. 'Reducing Existential Risks or Wild Animal Suffering? Part II: Expected Value Estimate'. The rational ethicist. https://stijnbruers.wordpress.com/2019/10/29/reducing-existential-risks-or-wild-animal-suffering-part-ii-expected-value-estimate/.

Centre for Effective Altruism. nd. 'History'. *Centre for Effective Altruism* (blog). nd. https://www.centreforeffectivealtruism.org/history.

Elga, Adam. 2010. 'Subjective Probabilities Should Be Sharp'. *Philosophers' Imprint* 10 (5): 1–11.

Gabriel, Iason, and Brian McElwee. 2019. 'Effective Altruism, Global Poverty, and Systemic Change'. In *Effective Altruism*, by Iason Gabriel and Brian McElwee, 99–114. Oxford University Press. https://doi.org/10.1093/oso/9780198841364.003.0007.

GiveWell. 2017. 'Deworming Might Have Huge Impact, but Might Have Close to Zero Impact'. The GiveWell Blog. https://blog.givewell.org/2016/07/26/deworming-might-huge-impact-might-close-zero-impact/.

Gracely, Edward J. 1996. 'On the Noncomparability of Judgments Made by Different Ethical Theories'. *Metaphilosophy* 27 (3): 327–32. https://doi.org/10.1111/j.1467-9973.1996.tb00212.x.

Greaves, Hilary. 2016. 'Cluelessness'. *Proceedings of the Aristotelian Society* 116 (3): 311–39. https://doi.org/10.1093/arisoc/aow018.

Greaves, Hilary, and Toby Ord. 2017. 'Moral Uncertainty About Population Axiology'. *Journal of Ethics and Social Philosophy* 12 (2): 135–67. https://doi.org/10.26556/jesp.v12i2.223.

Gustafsson, Johan E. forthcoming. 'Second Thoughts About My Favourite Theory'. *Pacific Philosophical Quarterly*.

Gustafsson, Johan E., and Olle Torpman. 2014. 'In Defence of My Favourite Theory'. *Pacific Philosophical Quarterly* 95 (2): 159–74. https://doi.org/10.1111/papq.12022.

Harman, Elizabeth. 2015. 'The Irrelevance of Moral Uncertainty'. In *Oxford Studies in Metaethics, Volume 10*, edited by Russ Shafer-Landau, 53–79. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198738695.003.0003.

Hedden, Brian. 2016. 'Does MITE Make Right?' In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, 102–28. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198784647.003.0005.

Herlitz, Anders. 2019. 'Cluelessness and Rational Choice: The Case of Effective Altruism'. *Ethical Perspectives*, no. 3: 401–26. https://doi.org/10.2143/EP.26.3.3287340.

Jackson, Frank. 1991. 'Decision-Theoretic Consequentialism and the Nearest and Dearest Objection'. *Ethics* 101 (3): 461–82. https://doi.org/10.1086/293312.

Joyce, James M. 2010. 'A Defense of Imprecise Credences in Inference and Decision Making'. *Philosophical Perspectives* 24 (1): 281–323. https://doi.org/10.1111/j.1520-8583.2010.00194.x.

Karnofsky, Holden. 2016. 'Hits-Based Giving'. Open Philanthropy. https://www.openphilanthropy.org/blog/hits-based-giving.

Lichtenberg, Judith. 2015. 'Peter Singer's Extremely Altruistic Heirs'. *The New Republic*, 2015. https://newrepublic.com/article/124690/peter-singers-extremely-altruistic-heirs.

MacAskill, William. 2014. *Normative Uncertainty*. Dissertation, University of Oxford.

———. 2016a. 'Normative Uncertainty as a Voting Problem'. *Mind* 125 (500): 967–1004.

———. 2016b. 'Smokers, Psychos, and Decision-Theoretic Uncertainty'. *Journal of Philosophy* 113 (9): 425–45. https://doi.org/10.5840/jphil2016113929.

———. 2018. 'Understanding Effective Altruism and Its Challenges'. In *The Palgrave Handbook of Philosophy and Public Policy*, edited by David Boonin, 441–53. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-93907-0_34.

———. 2019. 'The Definition of Effective Altruism'. In *Effective Altruism*, by William MacAskill, 10–28. Oxford University Press. https://doi.org/10.1093/oso/9780198841364.003.0001.

MacAskill, William, Krister Bykvist, and Toby Ord, eds. 2020. *Moral Uncertainty*. New product. New York: Oxford University Press.

MacAskill, William, and Hilary Greaves. 2021. 'The Case for Strong Longtermism'. *Global Priorities Institute, Working Paper No. 5-2021*.

MacAskill, William, and Toby Ord. 2020. 'Why Maximize Expected Choice-Worthiness?' *Noûs* 54 (2): 327–53. https://doi.org/10.1111/nous.12264.

MacAskill, William, and Theron Pummer. 2020. 'Effective Altruism'. In *International Encyclopedia of Ethics*, 1st ed., 1–9. Wiley. https://doi.org/10.1002/9781444367072.wbiee883.

MacAskill, William, Aron Vallinder, Caspar Oesterheld, Carl Shulman, and Johannes Treutlein. 2021. 'The Evidentialist's Wager'. *The Journal of Philosophy* 118 (6): 320–42. https://doi.org/10.5840/jphil2021118622.

Mogensen, Andreas L. 2021. 'Maximal Cluelessness'. *The Philosophical Quarterly* 71 (1): 141–62. https://doi.org/10.1093/pq/pqaa021.

Mogensen, Andreas L, and David Thorstad. 2020. 'Tough Enough? Robust Satisficing as a Decision Norm for Long-Term Policy Analysis.' *Global Priorities Institute Working Paper No. 15-2020*.

Nozick, Robert. 1995. *The Nature of Rationality*. Princeton (N.J.): Princeton University Press.

Podgorski, Abelard. 2020. 'Normative Uncertainty and the Dependence Problem'. *Mind* 129 (513): 43–70. https://doi.org/10.1093/mind/fzz048.

Resnik, Michael D. 1987. *Choices: An Introduction to Decision Theory*. Minneapolis: University of Minnesota Press.

Riedener, Stefan. 2021. *Uncertain Values: An Axiomatic Approach to Axiological Uncertainty*. De Gruyter. https://doi.org/10.1515/9783110736199.

Rinard, Susanna. 2015. 'A Decision Theory for Imprecise Probabilities'. *Philosophers' Imprint* 15 (7): 1–16.

Ross, Jacob. 2006. 'Rejecting Ethical Deflationism'. *Ethics* 116 (4): 742–68. https://doi.org/10.1086/505234.

Sebo, Jeff, and Laurie Paul. 2019. 'Effective Altruism and Transformative Experience'. In *Effective Altruism*, by Jeff Sebo and Laurie Paul, 53–68. Oxford University Press. https://doi.org/10.1093/oso/9780198841364.003.0004.

Sen, Amartya. 2017. *Collective Choice and Social Welfare: An Expanded Edition*. Harvard University Press. https://doi.org/10.4159/9780674974616.

Sepielli, Andrew. 2014. 'What to Do When You Don't Know What to Do When You Don't Know What to Do…'. *Noûs* 48 (3): 521–44. https://doi.org/10.1111/nous.12010.

———. 2017. *How Moral Uncertaintism Can Be Both True and Interesting*. Vol. 1. Oxford University Press. https://doi.org/10.1093/oso/9780198808930.003.0006.

Singer, Peter. 1972. *Famine, Affluence, and Morality*. Oxford ; New York: Oxford University Press.

———. 2015. *The Most Good You Can Do: How Effective Altruism Is Changing Ideas about Living Ethically*. Castle Lectures in Ethics, Politics, and Economics. New Haven ; London: Yale University Press.

Snowden, James. 2019. 'Should We Give to More Than One Charity?' In *Effective Altruism*, by James Snowden, 69–79. Oxford University Press. https://doi.org/10.1093/oso/9780198841364.003.0005.

Tarsney, Christian. 2018. 'Intertheoretic Value Comparison: A Modest Proposal'. *Journal of Moral Philosophy* 15 (3): 324–44. https://doi.org/10.1163/17455243-20170013.

———. 2019. 'Metanormative Regress: An Escape Plan'. *Working Paper.*

———. 2020. 'The Epistemic Challenge to Longtermism'. *Global Priorities Institute, Working Paper No. 10-2020.*

Todd, Benjamin. 2021a. 'Expected Value: How Can We Know What Makes a Difference When Uncertain?' 80,000 Hours. https://80000hours.org/articles/expected-value/.

———. 2021b. 'Is Effective Altruism Growing? An Update on the Stock of Funding vs. People'. 80,000 Hours. https://80000hours.org/2021/07/effective-altruism-growing/.

———. 2021c. 'What Is Social Impact? A Definition'. 80,000 Hours. https://80000hours.org/articles/what-is-social-impact-definition/.

Tomasik, Brian. 2014. 'Does Vegetarianism Make a Difference?' Reducing Suffering. https://reducing-suffering.org/does-vegetarianism-make-a-difference/.

Trammell, Philip. 2021. 'Fixed-Point Solutions to the Regress Problem in Normative Uncertainty'. *Synthese* 198 (2): 1177–99. https://doi.org/10.1007/s11229-019-02098-9.

Weatherson, Brian. 1998. 'Decision Making with Imprecise Probabilities'. http://brian.weatherson.org/vdt.pdf.

———. 2014. 'Running Risks Morally'. *Philosophical Studies* 167 (1): 141–63. https://doi.org/10.1007/s11098-013-0227-2.

———. 2019. *Normative Externalism*. First edition. Oxford, United Kingdom: Oxford University Press.

Wedgwood, Ralph. 2013. 'Akrasia and Uncertainty'. *Organon F: Medzinárodný Časopis Pre Analytickú Filozofiu* 20 (4): 483–505.

———. 2019. 'Moral Disagreement and Inexcusable Irrationality'. *American Philosophical Quarterly* 56 (1): 97–108.