Canadian Journal of Philosophy (2023), 53: 1, 1–20
doi:10.1017/can.2023.18

ARTICLE

There Is No Such Thing as Expected Moral Choice-Worthiness*

Nicolas Côté

Philosophy department, University of Glasgow, Glasgow, Scotland
Email: Nicolas.Cote@glasgow.ca.uk

Abstract
This paper presents some impossibility results for certain views about what you should do when you are uncertain about which moral theory is true. I show that under reasonable and extremely minimal ways of defining what a moral theory is, it follows that the concept of expected moral choiceworthiness is undefined, and more generally that any theory of decision- making under moral uncertainty must generate pathological results.

Keywords: moral uncertainty; normative ethics; probability theory; decision theory

1. Introduction

I frequently experience doubt about what is right and wrong, and sometimes about whether there is even such a thing as right and wrong. In recent years, numerous philosophers have argued that the way to deal with this doubt is to exploit the tools of decision theory (MacAskill 2013; MacAskill, Bykvist, and Ord 2020; King 2022). We can construe our moral uncertainty as uncertainty about which first-order moral theory is true. Each theory, in effect, represents a way the world might be, morally speaking, and so our moral uncertainty about the way the world is can be represented as a probability distribution (or perhaps a family of such distributions) over moral theories. Each moral theory, in addition, has a view about what is right and wrong in any situation—an opinion about which acts are more choice-worthy. Provided that for every theory we can assign numbers to acts representing (cardinally) how morally choice-worthy that act is according to the theory in question, and provided that these numbers can be meaningfully compared across theories, we can then convert the scary existential question of "what to do?" into a bland bureaucratic exercise in computing expectations of moral choice-worthiness and simply do what we expect will be morally best. Here's what this is supposed to look like:

Trolley Problem	Utilitarianism (5%)	Deontology (85%)	Virtue Theory (10%)
Push the large man	1000	–32	–11
Don't push him	–1000	15	15

In this table, the entries in the left column are the acts available to the individual, the entries in the top row are our objects of belief—moral theories—each of which has a subjective probability attached to it representing the individual's degree of belief in that theory, and the numbers in the

*Article last updated 25 October 2023.

boxes represent units of moral choice-worthiness. Under this representation moral theories are the "events" of standard decision theory—i.e., subsets of some underlying set, called the 'sample space,' and in our current example utilitarianism, deontology, and virtue theory are implicitly assumed to form a *partition* on this underlying sample space (this is important, for otherwise the subjective probabilities wouldn't sum to 1). The example I've chosen offers what appears to be a stark illustration of the power of decision theory to cut through the noise of moral disagreement: although the individual is quite confident that utilitarianism is false, and has credence 1 that if utilitarianism is false they should not push the large man in front of the trolley, the act which maximizes expected choice-worthiness is nonetheless to push the large man on the trolley tracks. If utilitarianism is true the world would be *so* much better if you pushed the large man and *so* much worse if you didn't that the value of these events swamps the implausibility of the theory, and therefore from your own point of view it's better to hedge your bets.

Much of the debate in the literature on moral uncertainty has focused on two big questions. First, are intertheoretic comparisons of moral choice-worthiness possible?[1] In other words, is it really possible to compare how desirable an act is according to different theories such that we can meaningfully assign numbers to those acts representing their desirability that can then be added up and combined in probabilistic mixtures? Or is taking a weighted sum of x's choice-worthiness score according to utilitarianism and its choice-worthiness score according to deontology similar to taking the sum of Mt. Fuji's altitude and my body temperature? And second, supposing that intertheoretic comparisons are possible, which decision rule should we adopt? That is, should we really maximize expected choice-worthiness, as Lockhart (2000), Enoch (2013), and MacAskill and Ord (2018) argue? Or should we instead simply do what the theory we're most confident in advises, as Gustafsson and Torpman (2014) argue? Or, finally, should we think, like Harman (2015) and Weatherson (2021, chap. 3), that our moral uncertainty is irrelevant, and that we ought to do just what the true moral theory says, and tough luck if we can't figure out what that is?

I am going to skirt these debates. Instead, I wish to raise a more fundamental problem. Philosophers of moral uncertainty tend to take for granted that we can speak meaningfully about degrees of belief in moral theories, yet rarely offer a formal definition of the probability space with respect to which these degrees of belief are defined.[2] A probability space is a mathematical object which assigns real numbers to events, here interpreted to represent degrees of belief in the truth of the event. The specification of the probability space is usually the first step in setting up a decision problem, and for good reason: probabilistic talk is only meaningful with respect to a probability space, which are richly structured objects that come with certain nonnegotiable properties that don't always gel with intuition. It seems to me that we are owed an explanation as to what, *precisely*, the objects of our belief are supposed to be when we are morally uncertain. Otherwise we're within our rights to doubt whether the manipulations being performed on these objects are at all meaningful: if you want to play with the tools of decision theory, you need to play by the rules of decision theory.

In this paper, I make two attempts at specifying the probability space of moral theories, and I show that for both these approaches we reach pathological results. Since the existing literature presupposes that moral theories can be represented as the events in a probability space, the first approach I discuss tries to specify moral theories as events. Catastrophic results emerge: among other problems, it turns out that this way of doing things implies that every act is necessarily permissible. Accordingly, the second approach I discuss tries to specify moral theories as possible outcomes. This approach does a bit better, but it instantly breaks down when you include as possible outcomes theories which can't be represented as orderings. The reason for this is simple: the general

[1]For the main protagonists in this debate, see Lockhart (2000), Ross (2006), Sepielli (2009; 2013), Gracely (2013), Hedden (2013), Hicks (2021), MacAskill, Bykvist, and Ord (2020), and Tarnsey (2021).

[2]The only exceptions I could find are Riedener (2020, 2021), Dietrich and Jabarian (2021), and Kernohan (2021). See section 3 for discussion.

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

idea behind moral uncertaintism is that in the event that some theory is true, an act is as desirable as that theory says it is. However, theories which can't be represented as continuous orderings, which includes many popular theories, do not say and cannot be made to say anything about how desirable any act is. These results lead me to doubt that there is a coherent concept of "expected moral choice-worthiness." And, more generally, they cast doubt on whether the moral uncertaintist approach, which consists of modelling moral doubt as uncertainty about the truth of "moral theories," is fruitful. This is because in some way or other theories are going to have to be modelled either as events or as outcomes, so the failure of my attempts to model them in either of these two ways does not bode well.

2. A primer on probability spaces

The project of moral uncertaintism only gets off the ground if individual beliefs can be represented by a probability distribution of some kind. So what is a probability distribution? Simply put, it's a numerical representation of uncertainty about what will happen. Let Ω be some arbitrary nonempty set of elementary outcomes; in a simple model of a dice toss, the set of outcomes might be 1,2,3,4,5, and 6. Next, define a family $\mathcal{F}$ of subsets of Ω such that $\mathcal{F}$ is closed under complement, countable union, and countable intersection: if A, B are both elements of $\mathcal{F}$, then $A \cup B$, $A \cap B$, and $B \backslash A$ are as well. $\mathcal{F}$ is what is called a σ-algebra over Ω. $\mathcal{F}$ is the set of possible events, i.e., propositions that might be true or false and that you could bet on; in our dice-toss model, the events will include "1," "2," etc., as well as "even,", "odd," "less than six," etc. Finally define a function $p : \mathcal{F} \rightarrow [0,1]$. p is a probability distribution if and only if it satisfies three conditions:

1. (Nonnegativity) $p(E) \geq 0 \, \forall \, E \in \mathcal{F}$.
2. (Unitarity) $p(\Omega) = 1$.
3. (Countable additivity) for any countable sequence $\{A_n\}_{n=1}^{\infty}$ of pairwise disjoint sets in $\mathcal{F}$, $p\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} p(A_n)$.

The triple $\langle \Omega, \mathcal{F}, p \rangle$ is called a probability space; Ω and $\mathcal{F}$ are called, respectively, the sample space and the event space. I shall interpret probabilities subjectively, as representing degrees of belief. Thus, the elements of $\mathcal{F}$ are our objects of belief. In our dice-toss model, if you think the dice is fair, then $p(\text{"even"}) = p(\text{"odd"})$.

If we want to use this sort of framework to represent individual beliefs about moral theories, we have no choice: either moral theories will have to be represented as events (i.e., elements of $\mathcal{F}$), or as sampling outcomes (i.e., elements of Ω). This is not an indifferent modelling choice: each entails a very different way of thinking about how moral theories are related to each other, and how our beliefs about them are structured. Under the first modelling choice, the objects of our beliefs are moral theories themselves and, formally, they are subsets of some common underlying sample space. Much more importantly, it means that the set of moral theories is a σ-algebra, and it therefore comes with nonnegotiable structure and content. Moral theories must bear set-theoretic relations to each other (e.g., some moral theories will be subsets of others), and all sets which can be obtained by countable intersection or union or complement from other moral theories themselves feature among the set of moral theories—and we must have beliefs about the moral theories thus obtained. Notice that under this first modelling choice multiple moral theories may be true: if A is a subset of B then $p(A) \leq p(B)$, and thus the truth of any moral theory implies the truth of any other theory of which it is a subset. Furthermore, if moral theories are events, then a number of principles of rationality can kick in which regulate how we must distribute our beliefs among them. For example, so-called regularity principles will require us to assign a nonzero credence to every nonempty event—i.e., every moral theory that isn't the empty set.

Under the second modelling choice, the objects of our beliefs are **not** moral theories themselves, but rather **sets** of moral theories, interpreted as events in which one of the moral theories in the set is

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

true. So for example, the event $\{T_1, T_2\}$ is the event in which either theory 1 or theory 2 is true, and the "atomic" event $\{T_1\}$ is the event in which T_1 is true. On this modelling approach, moral theories need bear no set-theoretic relations to one another, but they are all mutually exclusive. By definition, $p(A) = 1$ implies $p(\Omega \setminus A) = 0$. Thus, if moral theories are the elementary outcomes of our sample space, then no two theories can both be true, as all atomic events are pairwise disjoint. But most significantly, because the sample space is not a σ-algebra, the space of moral theories does *not* come with nonnegotiable content and structure: we can include what we like in our sample space, and neither probability theory nor EU theory can force us to include any particular theory in the sample space. By analogy, in our dice-toss example, the only outcomes I included were 1,2,3,4,5, and 6—I didn't also include the possible outcome "the dice vanishes in a puff of smoke"; I certainly could have included this possible outcome, but the laws of probability did not require it. This may turn out to be a blessing and a curse, as we will see.

The way in which the moral uncertainty framework is usually deployed suggests that moral theories are being thought of as events. The table in the introduction, which is typical of the literature, suggests that theories are supposed to be our objects of belief, and therefore our events. And it is widely assumed that theories can be fine or coarse grained arbitrarily: MacAskill and Ord (2018), for example, claim that we can always equivalently represent your credences between theories as being split 0.4-0.6 between deontology and utilitarianism, or instead as being split 0.4-0.3-0.3 between deontology and two slightly different, more specific versions of utilitarianism that disagree with one another but which both imply the more general version. This assumption provides the basis for an important argument against the 'My Favourite Theory' view (MFT), which holds that what you ought to do is just what "the" theory you find most plausible: no such thing, they claim, because the theory that gets the highest credence isn't the same depending on your choice of partition. Whether or not this argument is persuasive, it presupposes a model of theories as events, since some theories can imply others.

This immediately raises the question: What kinds of events?

3. Theories as events

I wish to consider three different ways of modelling theories as events: as sets of sentences, as preorders, and as choice functions. All three fail for a general reason, which is that on this approach, moral theories are "opinionated events," and the set-theoretic relations of different events force pathological combinations of opinions. Accordingly, the reader is invited to read just one of the following three subsections, then skip to section 4. For although the impossibility results shown throughout these subsections vary in detail, their inner logic is much the same.

3.a Theories = sets of sentences

One classic way of defining a moral theory is as a set of sentences—specifically, a set of sentences expressing moral propositions. We can make this idea precise by defining the object language $\mathcal{L} = \langle \mathcal{X}, \neg, \wedge, \vee, \bot, T \rangle$, where A is the set of possible actions and $\mathcal{X}$ is the closure of the set of all propositions of the form $\Diamond(x, S)$ where $x \in S \subseteq A$ under the unary operation $\neg$ as well as the binary operations $\wedge$ and $\vee$. The proposition $\Diamond(x, S)$ is interpreted to mean "it is permissible to choose x out of the set of options S. $\mathcal{X}$, in other words, is the closure under negation, conjunction, and disjunction of all simple deontic permission claims. For simplicity, I will denote elements of $\mathcal{X}$ by the generic terms $a, b, c\ldots$ I assume that $\mathcal{L}$ is a boolean algebra (see definition in appendix A).

A moral theory is any subset of $\mathcal{X}$, but of course many subsets of $\mathcal{X}$ represent completely pathological theories, including self-contradictory ones. It may therefore be useful to have a concept of a 'well-formed,' nonpathological theory in hand. The formal definition is in appendix A, but, intuitively, a well-formed moral theory $\mathcal{A}$ is a self-consistent set of deontic claims that is closed under implication, and that is maximally opinionated and nondilemmatic in the sense that for any

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

set of possible acts *S*, the theory tells you which of the available alternatives in *S* are permissibly chosen.

With that in mind, we can define the probability space $\langle \mathcal{X}, \mathcal{F}, p \rangle$, where $\mathcal{F}$ is some σ-algebra over $\mathcal{X}$ that includes all well-formed moral theories. For every moral theory $\mathcal{A}$, we define an arbitrary choice-worthiness function $f_{\mathcal{A}} : A \rightarrow \mathbb{R}$.

Now let us consider some basic adequacy conditions on a theory of decision-making under moral uncertainty.

V1—Valuation 1: for any $x \in S \subseteq A$, $\Diamond(x, S) \in \mathcal{A}$ implies $f_{\mathcal{A}}(x) \geq f_{\mathcal{A}}(y)$ for every $y \in S$. And $\neg\Diamond(x, S)$ implies $\exists y \in S : f(y) > f(x)$.

This is a very natural adaptation of the core claim of moral uncertaintism to the present modelling framework. If a theory says x is permissibly chosen out of the set of options *S*, then, supposing this theory is true, x must be at least as choice-worthy as every alternative in that set. And, in contrast, if a theory says x is not permissibly chosen out of the set of options *S*, then, supposing this theory is true, there must be some alternative in *S* which is strictly more choice-worthy. In effect, V1 is an analogue for value of David Lewis's (1980) principal principle, which states that you should believe that *P*, conditional on the chance of *P* being *p*, to degree *p*.

Some philosophers critical of moral uncertaintism have criticised this principle of valuation. For example, Weatherson (2014) argues that anyone who complies with this principle must have a fetishistic motivation: they must be motivated to act solely by a desire to do what's right (whatever that is), rather than by a desire to pursue those particular features of actions that make acts right. And it is morally vicious to do what's right merely because you believe it is right rather than because you believe it is good for others, or respects their rights, improves their character, or whatever. Likewise, Nissan-Rozen (2018) argues that anyone who specifically follows the rule of maximizing expected moral choice-worthiness must be motivated solely by a desire to maximize moral choice-worthiness, and not any concern for the first-order properties of acts that make them valuable.

For my purposes, it is not necessary to take a stance on whether adhering to V1 involves having fetishistic motivations. Perhaps it does. If so, we have a powerful external critique of moral uncertaintism. I wish to argue that even if we take on board the core assumption of moral uncertaintism, here expressed by V1, the project ultimately fails on its own terms. External worries about V1 can thus be set aside for now (though for a reply to Weatherson, see Sepielli 2009).

M1—MEC is defined 1: expected choice-worthiness is defined for at least one act, i.e., there exists a partition $\mathcal{A}_1, \ldots, \mathcal{A}_n$ of $\mathcal{X}$, such that $\mathcal{A}_i \in \mathcal{F}$, and some $x \in A$ such that $\sum_{i=1}^{n} p(\mathcal{A}_i) f_{\mathcal{A}}(x)$ is a real number.

This condition requires no special commentary. Many moral uncertaintists have argued that we should Maximize Expected Choice-worthiness (MEC). It is the very least one can ask of MEC that there exist at least one case where MEC expresses a meaningful claim.

NSI—Nonnecessity of Self-Inconsistency: it is not necessary that a morally uncertain agent believe with certainty that a logically impossible event is true, i.e., there is no $\mathcal{A} \in \mathcal{F}$ such that $p(\mathcal{A}) = 1$ and $\wedge_{a \in \mathcal{A}} \vDash \bot$.

Theorem 3.1. *V1 and M1 are inconsistent.*

Proof. Trivial. The proposition $\neg\Diamond(x, \{x\})$ is an element of $\mathcal{X}$ therefore by definition of a partition there must be some $\mathcal{A}_i$ such that $\neg\Diamond(x, \{x\}) \in \mathcal{A}_i$. By V1 it follows that $f_{\mathcal{A}_i}(x) < f_{\mathcal{A}_i}(x)$, which is impossible. Contradiction. □

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

It might be noted in this proof that $\mathcal{A}_i$ is not a well-formed theory. How is this consistent with my assumption that $\mathcal{F}$ included all well-formed theories? Simple: just because $\mathcal{F}$ includes *all* well-formed theories does not imply it *only* includes well-formed theories. Indeed, what the above proof shows is that it cannot only include well-formed theories. Because $\mathcal{F}$ is a σ-algebra on $\mathcal{X}$, the union of all its subsets is identical to $\mathcal{X}$. By definition, then, $\mathcal{F}$ must include a subset of $\mathcal{X}$ of which propositions like $\neg\Diamond(x,\{x\})$ is an element.

It might also be objected that $\neg\Diamond(x,\{x\})$ is a pretty pathological atomic moral proposition. Ought implies can, does it not? How can it fail to be permissible to do the only thing which you can? Fair enough, you can always choose a different sample space which excludes propositions like $\neg\Diamond(x,\{x\})$, in which case theorem 3.1 doesn't go through. But as it happens, many philosophers deny that ought implies can, for example because they believe in the possibility of moral dilemmas. And it's not as though $\neg\Diamond(x,\{x\})$ is a self-contradictory statement, so it's not clear to me why it would be wrong for an individual to entertain attitudes towards this proposition. And if the individual does entertain any attitude to propositions like $\neg\Diamond(x,\{x\})$, even if it's the attitude which assigns probability 0 to any theory which implies $\neg\Diamond(x,\{x\})$, theorem 3.1. applies. In any case, the following result cannot be escaped, even by choosing a different sample space.

Theorem 3.2. *NSI is a contradiction.*

Proof. Trivial. By unitarity, $p(\mathcal{X}) = 1$ and by definition $\mathcal{X}$ includes all propositions of the form $\Diamond(x,S)$ and all propositions of the form $\neg\Diamond(x,S)$. Thus, $\wedge_{a\in\mathcal{X}}\vDash\bot$. □

This result cannot be avoided even by restricting the outcome space to a subset of $\mathcal{X}$—not so long as one wishes to allow individuals to assign credences to more than one well-formed theory. So long as $\mathcal{F}$ includes at least two well-formed theories, the outcome space—which is the union of all subsets in $\mathcal{F}$—must necessarily include contradictory atoms for the simple reason that any two distinct well-formed theories disagree with each other. Only by restricting the outcome space to a subset of $\mathcal{X}$ which is itself a well-formed theory can you escape theorem 3.2. But of course to impose such a domain restriction is just to assume that the individual is certain about which moral theory is true, which moots the whole project of moral uncertaintism.

In other words, if we define moral theories as sets of sentences expressing moral propositions, and moreover we take theories to be the objects of our belief (i.e., events), then ipso facto we require morally uncertain agents to have credence 1 in contradictions, and moreover there is no act for which it is possible to define expectations of choice-worthiness. Note of course that this does not give us a reason to prefer MFT over MEC, because the failure of NSI is independent of our choice of decision rule. This modelling approach is therefore a complete flop.

3.b Theories = sets of preorders

So this first attempt has failed. Can we try anything else? Well, in many cases, the prescriptions issued by a moral theory to a given individual can be represented as a *preorder*, indexed to that individual on a given set of acts. A preorder on a set A is a binary transitive and reflexive relation on A, i.e., a subset R of the Cartesian product $A\times A$ with the property that $(a,a)\in R$ for all $a\in A$ and that if $(a,b)\in R,(b,c)\in R$ then $(a,c)\in R$. Nearly any moral theory can be represented as a preorder because provided that a given moral theory always permits at least one act to be chosen out of any given set of alternatives, and satisfies two mild consistency conditions,[3] we can then define R_T as the

[3]Condition α: T never prohibits choosing some act a out of a set of alernatives $B\subseteq A$ yet permits choosing a out of a larger set of alternatives $C\supseteq B$; and condition γ: T always permits some act a to be chosen out of a set of alternatives that is the union of a collection of sets of alternatives, each of which a is permissibly chosen from, according to T.

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

preorder such that $(a,b) \in R_T$ iff for any $S \subseteq A$, T permits a to be chosen from S if it permits b to be chosen from S. The choice function $\overline{T}$ induced by R_T will always agree with T about which acts are permissibly chosen out of any given set.[4]

Conveniently, preorders are already subsets of an underlying set of primitives, $A \times A$. Therefore, since moral uncertaintists want to think of theories as events, a natural choice of event space for the moral uncertaintist to take as the set of objects of some individual i's beliefs when i is morally uncertain is the smallest σ-algebra on $A \times A$ to include all i-indexed preorders on A. It turns out that there is only one such σ-algebra, which happens to be the same for all individuals—namely, the set of all relations on A that are either reflexive or irreflexive. This is easy to show (see appendix C). Let $\mathcal{F}$ denote this set.

So suppose we take $A \times A$ as our sample space and $\mathcal{F}$ as our event space, and then define for each individual $i \in \mathbb{N}$ a probability distribution $p_i : \mathcal{F} \rightarrow [0,1]$ representing their degrees of beliefs over the elements of $\mathcal{F}$, interpreted as moral theories. Our probability space is the triple $\langle A \times A, \mathcal{F}, p_i \rangle$; there exists one such probability space for each $i \in \mathbb{R}$. Does this construction make any sense? At first blush, you might think so; every preorder is included as an object of belief, so every rationally coherent moral theory finds representation in this framework, and even moral theories which violate rationality norms—such as moral theories which allow for intransitivities in value (Temkin 2012) or genuine moral dillemmas in which no act is permissible (Tessman 2015)—find representation, since at least some of these can be represented by cyclical relations, many of which are elements of $\mathcal{F}$.

Finally, define for every $R \in \mathcal{F}$ a function $f_R : A \rightarrow \mathbb{R}$ representing the moral choice-worthiness of acts according to R—equivalently, $f_R(x)$ is the choice-worthiness of x conditional on R. Now let's consider some reasonable conditions on a theory of choice under moral uncertainty.

V2—Valuation 2: $f_R(x) \geq f_R(y) \Leftrightarrow (x,y) \in R$.

V2 restates the key article of faith in the moral uncertainty program—namely, that you ought to desire some act x as much as moral theory T says x is desirable, supposing T is true. In fact, V2 is slightly weaker than this key article of faith because it only requires that if theory T is true and T instructs you to choose x out of $\{x,y\}$ then x must be morally more choice-worthy according to T than y.

N1—Nontriviality 1: it is *possible* for two acts $x,y \in A$ to be unequally choice-worthy, i.e., there is at least one x and y such that it is not the case that with probability 1, x, and y ar-e equally choice-worthy.

This nontriviality requirement is self-explanatory. If a theory of decision-making under moral uncertainty violates N1 and thereby implies that all acts are necessarily equally choice-worthy, it ipso facto reduces itself to absurdity. Whatever else is true about right and wrong, it is certainly not metaphysically, logically, or epistemically *necessary* that all acts are equally right and proper. A theory of decision making under moral uncertainty should not imply that any rational morally uncertain agent believes with certainty that every act is permissible. This is a datum. Besides, any theory implying the negation of this datum would strip itself of any action-guiding power, beating the point of this whole literature.

P—Possibility of wholesale moral error: it is *possible* that no substantive moral theory is true, i.e., there is an event E conditional on which no act is at least as choice-worthy as any other, and some individual i for which $p_i(E) \neq 0$.

[4]See Sen (1971).

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

Similar to N1, a theory of decision-making under moral uncertainty ideally should not require the truth of any substantive moral theory. If agents can be uncertain of which theory is true, they can be uncertain if any is true—it is open to them to doubt that moral predicates pick out anything real, to doubt that "good" "bad" "choice-worthy" are intensionally meaningful predicates. And indeed, many morally conscientious agents who worry about what is right and wrong also wonder whether there really is a right and wrong. Note that P does not *require* the individual to assign any positive credence to error theory. It just requires our decision theory to *permit* individuals to hold a positive credence in error theory. This shouldn't be too much to ask. The existential worries I've alluded to are common, especially among philosophers, and we should expect a fully general theory of decision-making under moral uncertainty to provide guidance to such morally conscientious agents.[5] The morally uncertain shouldn't have to check their logical positivst or antirealist worries at the door.

M2—MEC is defined: expected choice-worthiness is defined for at least one act, i.e., there exists a partition $R_1,\ldots,R_n$ of $A\times A$ and some $x\in A$ such that $\sum_{i=1}^{n}p(R_i)f_{R_i}(x)$ is a real number.

Theorem 3.3. *V2 and N1 are inconsistent.*

Proof. Recall that our probability space is $\langle A\times A,\mathcal{F},p_i\rangle$, for some $i\in\mathbb{N}$. $A\times A$ is our underlying outcome space. By the definition of a σ-algebra it is also a possible event. Indeed, it is the necessary event: by unitarity, $p(A\times A)=1$. By V2, $f_{A\times A}(x)=f_{A\times A}(y)\Leftrightarrow(x,y)\in A\times A$ and $(y,x)\in A\times A$. But $(x,y)\in A\times A$ for any $x,y\in A$, so $f_{A\times A}(x)=f_{A\times A}(y)$, violating N. □

Theorem 3.4. *V2 and P are inconsistent.*

Proof. Let $E\in\mathcal{F}$ be some event conditional on which no act is at least as choice-worthy as any other. Thus, there is no $(x,y)\in A\times A$ such that $(x,y)\in E$. Then $E=\varnothing$ and, necessarily $p_i(\varnothing)=0$, by unitarity and countable additivity. Our choice of i was arbitrary, therefore $p_i(E)=0$ for all i, violating P. □

Theorem 3.5. *V2 and M2 are inconsistent.*

Proof. Let $R_1,\ldots,R_n$ be some partition of $A\times A$, each $R_i\in\mathcal{F}$. (x,x) is an element of $\cup_{i=1}^{n}R_i$, so there is some R_i such that $(x,x)\in R_i$—call it R_1. All events in $\mathcal{F}$ are either reflexive or irreflexive, so R_1 is reflexive. Necessarily R_{2-n}, are all irreflexive: $(x,x)\in R,T$ for any x and any reflexive relations R,T. Irreflexive relations are not numerically representable, so if V2 holds, then $p(R_i)f_{R_i}(x)$ is undefined for all $i>1$. Thus, $\sum_{i=1}^{n}p(R_i)f_{R_i}(x)$ is undefined and not a real number, violating M2. □

One might be puzzled by the proof of theorem 3.3. Granted that, by the probability axioms, $A\times A$ is both the outcome space and the necessary event, and therefore the necessary moral theory, how can it be so opinionated as to regard all acts as equally good? Shouldn't it be maximally *un*opinioated, given that it is the union of all moral theories? To put it another way: if $A\times A$ is the necessary moral theory, shouldn't its assessments of chocieworthiness be consistent with the assessments of any particular moral theory, and therefore be empty? Well, no. That's what the theorem shows! This is indeed a strange result, but it is simply the pathological consequence of (1) modelling moral theories as events, (2) defining events as relations on a set of acts, and (3) requiring the choice-worthiness of

[5]Obviously, if error theory under the guise of nihilism is true, we can't provide *moral* guidance to the individual. But we may still be able to provide rational guidance to them, given their aims and beliefs.

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

an act conditional on some event E to numerically represent the ordering E induces on the set of acts. The simple fact is, the union of all relations on a set of acts is a relation on that set of acts. And not just any relation: it is the relation which R-relates every act to every other. If the union of all relations is the necessary event, then all acts must necessarily be R-related to every other act. And if all acts are R-related to one another by the necessary moral theory R, then all acts are equally good.

The reader may also complain that in the proof of theorem 3.5, R_{2-n} are not ranking relations—they're not well-formed theories, so how is it that they can be elements of $\mathcal{F}$? Again, this is a consequence of the fact that $\mathcal{F}$ is a σ-algebra: as explained above, the *smallest possible* σ-algebra on $A \times A$ to include only well-formed theories is the set of *all* relations on $A \times A$ that are either reflexive or irreflexive, and this set includes a number of relations on $A \times A$ that are not ranking relations.

In sum, under this definition of a moral theory, moral uncertaintism requires us to believe with certainty that everything is permitted, and with certainty that error theory is wrong. MEC is also undefined (though again, this is cold comfort for MFT, since the former results hold regardless of our choice of decision rule).

3.c Theories = sets of choice functions

So our second attempt was also a flop, but let's not give up yet. One last try. Suppose we try to define moral theories as choice functions instead of as preorders. If A is our set of possible acts, a choice function is a function which selects from every nonempty subset B of A a nonempty subset of B. Intuitively, a choice function selects from a menu of alternatives a submenu of alternatives, each of which is permissibly chosen. Formally, a choice function c is a subset of the Cartesian product $\mathcal{P}_0(A) \times \mathcal{P}_0(A)$, where $\mathcal{P}_0(A) = \mathcal{P}(A) \backslash \varnothing$, with the property that for any two nonempty $B, C \in A, (B, C) \in c \Longrightarrow C \subseteq B$.

Let us now define the probability space $(\mathcal{P}_0(A) \times \mathcal{P}_0(A), \mathcal{E}, p)$, where $\mathcal{E}$ is any σ-algebra over $\mathcal{P}_0(A) \times \mathcal{P}_0(A)$ that includes all possible choice functions over A. Unfortunately, since $\mathcal{E}$ is a σ-algebra, it cannot include only choice functions, it must also include as elements the complements of choice functions, and all countable unions and countable intersections of choice functions and their complements. And these other elements are not themselves choice functions. We need some way of associating choice-worthiness judgments with every element in $\mathcal{E}$. One way of doing this is to associate with every element in $\mathcal{E}$ a "pseudo" choice function that selects from every set of alternatives a subset of that set whose elements are permissibly chosen—but unlike with ordinary choice functions, this subset may be empty. If the pseudo choice function maps a set of alternatives to the empty set this means that nothing is permissibly chosen, as might be the case if we face a true moral dilemma. So for every $R \in \mathcal{E}$, define the pseudo choice function C_R as:

$$C_R = \{(B, C) \in \mathcal{P}(A) \times \mathcal{P}(A) | C \subseteq B \text{ and } \forall x \in C, x \in B \Leftrightarrow \exists D \in \mathcal{P}(A) : (C, D) \in R \text{ and } x \in D\}.$$

Under this definition, an event $R \in \mathcal{E}$ regards some act x as permissibly chosen out of a set of alternatives B if and only if B is R-related to another set D which contains x. This definition has the attractive property that if R is a choice function, then $C_R = R$. The intuition here is that, ordinarily, if some set A is R-related to a set B by a choice function R, we take this to mean that the elements of B are those acts which are permissibly chosen given that A is the set of available alternatives. Thus, if we want to interpret the elements of $\mathcal{E}$ as telling us something about what can be chosen, we should try to preserve the idea that if some set A is R-related to another set B by some relation $R \in \mathcal{E}$, then the elements B are acts which are permissibly chosen given that A is the set of alternatives. The trouble is that if $\mathcal{E}$ is not a choice function, it may R-relate A to several different sets, including sets which don't contain any elements from A! A conservative proposal for how to interpret the choice instructions R is giving us is then that if A is R-related to B, then *if* B contains elements from A, then those elements must be permissibly chosen from A. With this definition in hand, let's consider some reasonable conditions on a theory of decision-making under moral uncertainty.

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

V3—Valuation 3: $f_R(x) \geq f_R(y) \Leftrightarrow \forall B \subseteq A : x,y \in B, y \in R(B) x \in R(B)$.

N2—Nontriviality 2: it is possible for two acts x, y to be unequally choice-worthy, i.e., it is not the case that with probability 1 x and y are equally choice-worthy.

Theorem 3.6. *V3 and N2 are inconsistent.*

Proof. Trivial. By unitarity, $p(\mathcal{P}_0(A) \times \mathcal{P}_0(A)) = 1$, and by definition of the pseudo choice function $C_{\mathcal{P}_0(A) \times \mathcal{P}_0(A)}(B) = B$ for any $B \in \mathcal{P}_0(A)$, because all pairs of the form (B, C) are elements of $\mathcal{P}_0(A) \times \mathcal{P}_0(A)$. By V3 it then follows that $f_{\mathcal{P}_0(A) \times \mathcal{P}_0(A)}(x) = f_{\mathcal{P}_0(A) \times \mathcal{P}_0(A)}(y)$, which violates N2. □

Thus, all three attempts so far at modelling theories as events have foundered. The impossibility results we've gone through appear to be robust, which suggests that the difficulties with modelling theories as events is not sensitively tied to just what *kind* of event we choose to model them by. Rather, the problem seems to be with trying to be with trying to represent theories as events in the first place—specifically, with trying to represent them as "opinionated" events. And since the opinionation of events is a consequence of the fundamental commitment of moral uncertaintism (here expressed by V1–V3) to the claim that we ought to desire an act to the extent that a theory says the act is desirable, the problem cannot be avoided. In sum, it is doubtful that the project of moral uncertainty can succeed by representing theories as opinionated events.

4. Theories as possible outcomes

But if all our problems stem at least partly from the fact that we wanted to treat moral theories as events, the alternative is to treat them as outcomes. Our sample space is therefore some set Ω of well-formed theories, or choice functions, or preorders. Without loss of generality, let's say that Ω is a set of preorders, and let's take the full powerset $\mathcal{P}(\Omega)$ as our event space. Intuitively, we can interpret the event $\{R\}$ as the event in which the moral theory defined by R is true, and the event $\{R_1, R_2, \ldots, R_n\}$ as the event in which one of the theories defined by $R_1, \ldots, R_n$ is true. We can then define a probability distribution p_i over $\mathcal{P}(\Omega)$ representing individual i's beliefs about which moral theory is true.

This way of doing things involves a significant departure from the way moral uncertaintists want us to think about moral uncertainty. And in fact, the motivation for moral uncertaintism is sharply undermined if we have to represent theories as sampling outcomes. It is central to the motivation behind this entire literature that moral theories can be thought of as events: one of the key reasons defenders of MEC are attracted to it is that it is supposed to show that even anti-utilitarians who have no confidence in utilitarianism must nonetheless, as a requirement of rationality, substantially follow its directives. Here's how the argument is supposed to work. We should maximize expected choice-worthiness, according to MacAskill, Bykvist, and Ord (2020), because what you should prefer to do ought to be sensitive to what those theories you have at least *some* confidence in say you ought to do. Given that it is irrational—according to the so-called 'regularity' principle—to assign zero credence to nonempty events, you ought to have some confidence in every logically possible theory, including utilitarianism. And because utilitarianism has very strong opinions about what is right and wrong (the choice-worthiness values it assigns exhibit higher variance than those assigned by other theories), this entails that even if you don't have much credence in utilitarianism, you should heed its demands to a good degree. The badness (in utilitarianism's opinion) of not following utilitarianism's demands swamps its improbability.

I have my doubts that regularity really is rationally mandated, but let's grant that it is. In that case, the argument works, but *if and only if* moral theories are events! Since any subset of our sample space defines an event, and therefore a theory that you have no choice but to consider, the regularity

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

principle forces you to assign this theory a nonzero probability. This argument appears frighteningly strong; in fact: even the moral views expressed by Nazism will occupy some subset of our sample space, and Nazism is also very opinionated, and therefore regularity will require our own attitudes to be strongly positively responsive to the views of the Nazis.

But if moral theories are modelled as outcomes, the argument falls apart. Regularity is a requirement on the distribution of real numbers to events, not a requirement on our choice of sample space. Remember, the laws of probability don't require me to include the outcome "the dice vanishes in a puff of smoke" in my model of a random dice roll. And indeed, if you go to a casino, you will find that the odds they give you in Craps ignore the possibility of such outcomes. Therefore, if theories are outcomes, then rationality cannot force you to consider theories you would prefer to ignore. If I am attracted to moral uncertaintism as a mode of deliberation, but the only moral theories I want to consider are deontology and virtue ethics, rationality can't stop me. I can easily construct a well-behaved probability space with only two outcomes, whose only atomic events are therefore $\{R_{\text{deontology}}\}$ and $\{R_{\text{virtue ethics}}\}$, and these two events will jointly exhaust the space of possibilities in my model. Every nonempty event in my event space will receive a nonzero probability, satisfying the regularity principle, and yet utilitarian considerations do not feature in my deliberations. You can try to persuade me that it's somehow unwise to exclude a theory you like from my sample space (I'm skeptical: if a Nazi made that argument to me I wouldn't be persuaded to include Nazism in my sample space), but you can't show it's irrational. In sum, if you are attracted to moral uncertaintism because you find the regularity argument persuasive, then you can't allow moral theories to be represented as outcomes. Conversely, the fact that we have to represent theories as outcomes should disenchant you with moral uncertaintism. If moral theories are sampling outcomes, then the hope of persuading nonutilitarians to be utilitarians despite themselves fails.

Granted, some moral uncertaintists might reply that their efforts to convince nonutilitarians to be utilitarians despite themselves don't really depend on regularity. Rather, they might just assert that, as a matter of empirical fact, most people do find utilitarianism at least somewhat plausible and so do consider its claims in their deliberations. If that's the argument, then it's no great loss to represent theories as outcomes rather than events; but then again, if that's the argument, then the argument loses its force. If your efforts to persuade people to be utilitarians despite themselves depends on their already being disposed to act according to utilitarian recommendations, persuasion is either unnecessary (because they're already disposed to do what you're nagging them to do) or futile (because your persuasion strategy appeals to dispositions they lack). What was supposed to be impressive about these "hedging" arguments was their potential to force resolute anti-utilitarians to take utilitarianism more seriously despite being generally unwilling to consider its claims in their deliberations, simply by appealing to weak norms of rationality that stand independently of utilitarianism.

Likewise, the central argument, which we discussed earlier, by MacAskill and Ord (2018) against the My Favourite Theory view is in trouble. Recall that the thrust of the objection is that it doesn't make sense to talk of "the" theory you find most credible because we can always fine grain or coarse grain the level at which we describe theories, and so the theory which gets the highest credence depends on an arbitrary choice of partition of the state space. This objection, though, assumes that theories are modelled as events. If instead we model theories as possible outcomes, the objection has no force, because it suffers from a catastrophic presupposition failure: to say that R is your favourite theory is to say that the atomic event you give highest credence to is $\{R\}$. But remember that $\{R\}$ is not a moral theory—it is the atomic event in which some particular moral theory is true. Since it is an atom, it cannot be fine grained. You can of course coarse grain an atomic partition of the state space, but the events you end up with are not increasingly general moral theories, just increasingly disjunctive events in which one of many moral theories is true: the event $\{R_1, R_2\}$ is the event "either R_1 is true or R_2 is true," not some highly general moral theory which is implied by both of R_1 and R_2. Moral theories under this definition cannot be fine grained or coarse grained.

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

Note that the point I am making here is quite different to the point Gustafsson and Torpman (2014) make in response to MacAskill's argument: their response is to simply assume that there is a most natural partition—not necessarily the atomic partition—in terms of which the claims of MFT should be couched. The natural partition is something like the level of description of moral theories that you prefer, the one you find most salient. Your favourite theory is therefore the theory you prefer at the level of grain you find most salient. But note that this way of thinking about things necessarily models theories as events in an event space which can in principle be fine or coarse grained. If theories are sampling outcomes though, this whole response also suffers from presupposition failure.

As compensation for this departure, though, we avoid some of the absurdities of our earlier approach. It's no longer the case that the individual must assign probability 1 to it being true that every act is just as choice-worthy as any other. Instead, the individual must only believe with probability 1 that some theory (possibly including error theory) is true—a much more plausible conclusion. This is because the necessary event $\{R_1, R_2, \ldots\}$ is not itself a moral theory. In particular it is not the *union* of all preorders representing the moral theories under consideration. It is the *set* of all such preorders, an event which we interpret as the event in which one of the theories under consideration is true.

Further, the process of partitioning the sample space works just as you would hope: if Ω contains the ten moral theories you find yourself hesitating between as elements, then you can partition Ω into ten subsets, each containing a single element corresponding to one of those ten theories you're uncertain about. We now even have a way, it seems, of defining expectations of moral choice-worthiness. Suppose, for simplicity, that Ω is finite, and that every element of Ω is a complete preorder.[6] Then each element of Ω can be represented by a real-valued function f_R. From here, we define the expected value of an act x as $E(x) = \sum_{R \in \Omega} p_i(\{R\}) f_R(x)$. Thus, the choice-worthiness of an act is the probability-weighted sum of its choice-worthiness under each moral theory. Provided that every moral theory included as a possible outcome in the sample space is representable by a real-valued function, $E(x)$ is well-defined, and you can meaningfully instruct individuals to maximize expected choice-worthiness. Finally, the process of moral learning works roughly as you would hope: if you read the works of Peter Singer and become convinced that eating meat is wrong, then you've effectively "observed" the event in which at least one theory which implies that eating meat is wrong is true—thus by updating by conditionalization you come to assign credence 0 to the atomic events in which theories which fail to imply this are true, and you raise your credence in each atomic event in which one of the remaining theories is true.

This approach is a slighly more general version of that proposed by Riedener (2020) and Dietrich and Jabarian (2021). They model moral theories as vNM utility functions, then postulate a relation on the set of probability distributions over these functions (i.e., implicitly, the vNM functions are treated as sampling outcomes), which they assume also to satisfy the vNM axioms. As Dietrich and Jabarian helpfully explain, this model effectively imagines that individuals are behind a veil of ignorance, and don't know what their values will be. They then exploit results like Harsanyi's theorem to argue that the maximization of expected value is the only rule for aggregating the different vNM functions which satisfies the *ex ante* Pareto axiom (i.e., if every theory expects A to be better than B, then you should prefer A to B). The only difference between these approaches and the one I've just sketched is that I don't assume that moral theories must satisfy the vNM axioms, nor that any acceptable decision rule under uncertainty must also satisfy these axioms. It's worth noting that Riedene, Dietrich, and Jabarian are, to my knowledge, the only authors to provide an explicit probability model of moral uncertainty, so it may speak in favour of representing theories as sampling outcomes that this is how they themselves do it.

[6](If A is infinite, we just need to add a technical condition of continuity.)

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

Maybe all of this is compensation enough for heterodoxy. Unfortunately, while this setup may avoid the blatant absurdities of the previous approach, its modelling capacities are so limited that it fails as a general model of agents experiencing moral doubt, and it cannot successfully be used as a tool in deliberation by such agents either. This can be shown by another impossibility result. Define a choice rule for decision-making under moral uncertainty (or "choice rule," for short) as a map from a set of possible probability spaces $(\Omega, \mathcal{F}, q)$ to a set of preorders over A, such that Ω is a set of binary relations on A. Intuitively, given that you are uncertain between a given set of moral theories and your uncertainty over these theories is representable by a probability distribution, the choice rule tells you what to do. Now consider the following conditions.

U—Universal domain: every logically possible probability space $(\Omega, \mathcal{F}, q)$ is included in the choice rule's domain.

This is another way of saying that individuals are allowed to form doxastic attitudes to any moral theory. Individuals may not be *required* to consider any particular theory in their deliberations, as argued earlier, but surely they must be *permitted* to consider any moral theory they wish. Our theory of decision-making under uncertainty must be able to guide and advise any agent who is uncertain of which moral theories are true. If there is some moral theory in which they in fact express some confidence our choice rule must be able to guide them. A choice rule which only works when restricted to a narrow range of possible moral theories fails as a general solution to the problems of moral doubt. A slightly less permissive, but still quite permissive alternative to U is the following.

Q—Quasi-universal domain: every probability space $(\Omega, \mathcal{F}, q)$ such that Ω only includes reflexive relations is included in the choice rule's domain.

By strengthening U to Q, we prohibit the individual from taking doxastic attitudes to the empty set, but also to pathological moral theories that do claim that some acts are more morally choice-worthy than others while also denying that any act is at least as choice-worthy as itself. After all, maybe it's too much to ask a choice rule to provide sensible guidance when the individual has beliefs in moral theories that are simply unintelligible. But at the very least, a theory of decision-making under uncertainty should work for any agent who expresses doxastic attitudes to moral theories that are actively debated by moral philosophers. No such theories (other than error theory) are excluded by Q. Next we define MEC and MFT in this new modelling framework.

(MEC) Maximize Expected Choice-worthiness: for any $(\Omega, \mathcal{F}, q)$ in the domain of our choice rule, the image of $(\Omega, \mathcal{F}, q)$ under this choice rule is the preorder such that $x \cdot y \Leftrightarrow \sum_{R \in \Omega} q(\{R\}) f_R(x) \geq \sum_{R \in \Omega} q(\{R\}) f_R(y)$

(MFT) My Favourite Theory: for any $(\Omega, \mathcal{F}, q)$ in the domain of our choice rule, the image of $(\Omega, \mathcal{F}, q)$ under this choice rule is the preorder $\gtrsim$ such that $x \gtrsim y \Leftrightarrow x \overline{R} y$, where $\overline{R} \in \Omega$ is such that $q(R) \geq q(T)$ for any other $T \in \Omega$.

This condition can be extended to lexical order in case there are ties.

(LMFT) Leximin My Favourite Theory: for any $(\Omega, \mathcal{F}, q)$ in the domain of our choice rule, define a sequence $\{R\}_{i=1}^{|\Omega|}$ such that every $R_i \in \Omega$ and $q(R_n) \geq q(R_{n+1})$. The image of $(\Omega, \mathcal{F}, q)$ under this choice rule is the preorder $\gtrsim$ such that $x \succ y \Leftrightarrow \exists n \in \mathcal{N}$ such that for all r: $1 \leq r < n, x R_r y$ and $x R_n y$. Moreover $x \sim y \Leftrightarrow x R_i y$ and $y R_i x$ for all $R_i \in \Omega$.

Theorem 4.1. *There exists no choice rule satisfying Q and MEC.*

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

Proof. Q requires the choice rule to yield a preorder over A for a probability space $(\Omega, \mathcal{F}, q)$ such that Ω includes cyclical relations on A; Q also requires the choice rule to yield a preorder for Ω that includes lexicographic relations. But for any cyclical relation R f_R is undefined, as is f_R if R is a lexicographic relation over A and A is infinitely large, thus no preorder $\succsim$ exists satisfying the condition that $x \succsim y \Leftrightarrow \sum_{R \in \Omega} q(\{R\}) f_R(x) \geq \sum_{R \in \Omega} q(\{R\}) f_R(y)$. Thus, MEC is violated. □

Theorem 4.2. *There exists no choice rule satisfying U and MFT.*

Proof. U requires the choice rule to yield a preorder over A for a probability space $(\Omega, \mathcal{F}, q)$ such that Ω includes the empty set and $q(\varnothing) > q(R)$ for any $R \neq \varnothing : R \in \Omega$. No preorder $\succsim$ exists such $x \cdot y \Leftrightarrow (x, y) \in \varnothing$. Thus, if U is satisfied MFT is violated. □

Theorem 4.3. *There exists no choice rule satisfying U and LMFT.*

Proof. Follows directly from theorem 4.2. □

Note that theorems 4.2 and 4.3 hold if we replace U with any domain condition that requires the choice rule to yield a preorder for probability spaces for which the error theory is a possible sampling outcome. U is much stronger than we need.

How damaging are these results? One can certainly escape theorem 4.1 by imposing a more restrictive domain condition than Q. This is what Riedener (2020) and Dietrich and Jabarian (2021) do by assuming all moral theories satisfy the vNM axioms (including transitivity and continuity), and their approach delivers well defined expectations of moral choice-worthiness.[7] But, for my money, their domain assumption is so restrictive that it simply defeats the project of moral uncertaintism. They assume that the only moral theories you care to consider in your deliberations are ones that are already in the business of maximizing expected value. Therefore, if in the course of your moral deliberations, you wish to entertain and engage with anti-aggregative moral theories like contractualism, or 'limited aggregation' theories (Voorhoeve 2014; Steuwer 2021), or value theories like Larry Temkin's, which recognises value cycles, then your state of mind is not one which can be modelled on their approach. This is a problem.

For a start, it means that the Riedener-Dietrich-Jabarian (RDJ) approach offers no general answer to the question which moral uncertaintists were out to answer, namely: what should I do when I'm unsure what I should do? Moreoever, it consigns the RDJ approach to total irrelevance in the practical deliberations of anyone who wishes to engage with nonmaximizing moral theories in their moral deliberations. This will include many Bayesians and consequentialists who are otherwise sympathetic to value maximization. I, for example, am not a anti-aggregative deontologist, but I don't think Scanlonian contractualism is *insane*: I feel the pull of the TV transmitter room case, and I think that responsible moral inquiry must respond to contractualist arguments and thought experiments. The RDJ approach therefore cannot feature prominently in responsible moral inquiry, since it prohibits me from entertaining any attitudes to contractualism. This self-imposed irrelevance also forestalls any possibility of convincing anyone who wishes to engage reflectively with deontology to "hedge their bets" in favour of consequentialist moral theories. This is for the same reasons as above: any strategy for persuading me to be less of a deontologist that already assumes I don't consider any form of deontology relevant to my deliberations is a failure.

We need to be at least somewhat permissive in the range of moral theories we allow individuals to express attitudes to, else the project of moral uncertaintism fails on its own terms. Q, I think, is one of the most restrictive domain condition one can impose without vitiating the aims of moral

[7]Kernohan (2021) adopts a different approach in which theories are modelled by random variables, but this difference is insubstantial. Theories are still sampling outcomes and are defined as real-valued functions.

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

uncertaintism.[8] The inconsistency of Q with MEC is therefore, I submit, a challenge to MEC's plausibility.

That's it for theorem 4.1. What about theorems 4.2 and 4.3? Is it such a problem that MFT bottoms out of advice when individuals are unsure of whether error theory is true? Q and LMFT are consistent (thus, by extension, Q and MFT), so LMFT will provide sensible guidance for anyone who does believe there is a right and wrong, but is merely uncertain of what makes something right or wrong. This, I concede, is already pretty good. MFT is less severely impugned by the above results than MEC, but nonetheless I claim that MFT's inability to guide agents who are in doubt as to whether anything is right or wrong is a serious failing. Not only does MFT fail to provide guidance to all morally uncertain agents, and thus fails as a general theory of decision-making under moral uncertainty, this failure looks like a serious fault in the approach because, curiously, I don't find it particularly difficult to think of reasonable advice for such an agent. If they find error theory such a threatening possibility, they could be invited to test their reasons for finding it so plausible by reading a healthy dose of Putnam (2002, 2004), and Williamson (Forthcoming).

This last remark might seem a bit flippant, but it points to what I believe is a general and underappreciated flaw with the whole approach proposed in this literature. The rules that have been proposed for decision-making under moral uncertainty are, by design, completely insensitive to the preferences and nonmoral beliefs of individuals. This is the root of many problems. For a start, it leaves these theories unable to rationalize a range of ordinary morally conscientious behaviour. For example, a morally conscientious agent who is unsure what to do might find it quite desirable to spend time reading about and trying to settle metaethical debates, or engage in moral disputation to try and settle what is right and wrong, or seek advice about what to do and how to live from their moral gurus or heroes; these strategies stand a chance of resolving their underlying uncertainty (see e.g., Weatherson 2014; Sepielli 2016). Just as we expect a good theory of decision-making under ordinary descriptive uncertainty to have rationalize ordinary evidence-seeking behaviour, and rationalize our willingness to pay for information, we should expect a good theory of decision-making under moral uncertainty to rationalize the evidence-seeking behaviour I've described. But, in fact, it's very hard to see how the approach sketched out in this section could *ever* recommend the activity of moral inquiry, because no first-order ethical theory will ever advise such behaviour: from each theory's own point of view, all moral issues are settled, and you simply ought to do what the theory says and, consequently, MFT and MEC will also advise against such moral evidence-seeking.

Unless the moral theory finds intrinsic value in moral deliberation as such (this would be a strange theory indeed), or unless moral deliberation happens to have instrumentally valuable causal effects (e.g., it turns out that, for some arbitrary reason, an act of moral reflection will maximize welfare), any moral deliberation must necessarily compete for your time and energy with other courses of action that actively follow the theory's guidance, and for that reason moral deliberation will be disfavoured in comparison to those alternatives. To take a stylized example, suppose the individual is uncertain between theories $T_1, \ldots, T_n$, and has options $A_1, \ldots, A_{n+1}$ to choose from where A_i is the option which T_i regards as optimal and A_{n+1} is the option "take longer to reason about which moral theory is true." From the agent's own point of view, A_{n+1} might be the optimal course of action because taking the time to reason more might substantially resolve their moral uncertainty. But neither MFT nor MEC will recommend A_{n+1} because no particular moral theory will ever recommend A_{n+1}. So MFT and MEC can't rationalize ordinary evidence-seeking behaviour.

[8] I take this critique of MEC to be broadly complementary to Nissan-Rozen's (2015) critique, which is that even if we restrict ourselves to moral theories representable as real-valued functions, there is no decision rule which satisfies the basic adequacy conditions of state dominance and ex ante Pareto while being even minimally permissive about what kind of risk attitudes moral theories are allowed to have.

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

In fact, it's not even clear that these approaches have a complete concept of evidence. Most moral uncertainty-ists postulate agents who have no beliefs about anything other than which theory is true, and even those few who do explicitly model individual's descriptive uncertainty alongside their empirical uncertainty (e.g., Riedener [2020], Dietrich and Jabarian [2021], and Kernohan [2021]) do so in a way that makes it impossible to represent the dependence of moral theories on empirical beliefs.[9] The only way individuals can change their moral beliefs is by "observing" the truth of particular moral principles or even entire theories. But our reasons for favouring one ethical theory over another are frequently guided by beliefs about empirical matters.[10] For example, my skepticism of preference-satisfactionist utilitarianism is partly driven by my skepticism about empirical matters: I'm skeptical that preferences are cardinally measurable and interpersonally unit comparable. My skepticism could be abated if I were to learn of a new measurement procedure that allowed for truly cardinal measurements of preferences. But neither this learning event nor the change it would induce in my beliefs can be represented in a probabilistic model where the only objects of beliefs are moral theories, because neither my beliefs about measurement nor the evidence I've observed are about moral theories. (There's also a separate concern about what moral uncertaintists should say about what we should do in cases where, conditional on some empirical fact, some moral theory cannot be meaningfully stated, but I put this aside.)

Likewise, the advice we give to the morally conscientious should also depend on their preferences. If it just matters to them that *some* moral theory—*any* moral theory—be true because the prospect of nihilism is just too painful to contemplate—like Ivan Karamazov, our agent *needs* the tears of children to matter—then it seems suitable to advise them to consult defenses of moral realism, or perhaps read good existentialist novels like *The Plague* by Camus. In contrast, if your moral doubt and anxiety are driven by your concern for the poor yet simultaneous discomfort with coercing people by the (implicit) threat of bayonets into giving away their assets, and what you really want is either a way of making socialism safe for libertarianism (and vice versa), or a way of overcoming your lingering attachment to one or the other, then it seems more suitable to advise you to adopt a stance of cautious liberalism while you consult the back-and-forth between different stripes of libertarian, anarchist, liberal, and socialist political philosophers. The uncertainty we care most about resolving is not the same in these two cases, and good advice to the morally conscientious should be suitably responsive to this fact. But existing models of decision-making under moral uncertainty cannot be responsive to this fact since they are by design unresponsive to people's preferences.

This is all by way of saying that present approaches to decision-making under moral uncertainty are severely limited, and the problem we have here is fundamental to the modelling approach favoured by the moral uncertaintists. The above impossibility results show that if moral theories are to serve both as objects of belief and as constraints on the distribution of values whose expectation we are supposed to maximize, then we face a trade-off between choosing a measure space that is adequate to represent the space of possibilities which individuals are aware of/wish to consider in their deliberations, and one that does not impose constraints on the distribution of values to acts which are logically impossible to satisfy. And the brief remarks I've just offered suggest that they fail in other ways too, by being unresponsive to precisely those individual beliefs and concerns we should expect a good source of advice to be responsive to. Moral doubt is a poignant experience of

[9]This is because empirical propositions are modelled as sets of possible worlds and moral theories are not. In probability theory, evidential relations between propositions are modelled by set-theoretic relations between those propositions. But no one has ever proposed a model of moral uncertainty on which theories and empirical propositions bear set-theoretic relations to one another. It's hard to see how they could, as this would require modelling theories as events, and we know how that goes.

[10]Some philosophers have raised the contrary worry for skeptics of moral uncertaintism like myself that some of our empirical uncertainty can depend evidentially on our moral uncertainty; see e.g., MacAskill, Bykvist, and Ord (2020), Podgorski (2020), and Robinson (2021) for discussion. I am skeptical that these arguments succeed, largely again because the probability spaces in these arguments are never clearly defined, but I won't press this point.

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

life, but inventing a whole new decision theory featuring preferenceless agents who only have beliefs about a narrow class of objects to talk about it mires us in confusion.

5. Conclusion

There are many important questions that remain unsettled about what to do when you are in the throes of moral doubt. What I have argued here, however, is that it is not a fruitful way to try and address these questions by modelling moral doubt as a situation where we are uncertain over the truth "moral theories," construed as very special, very unusual sorts of events or outcomes in a probability space. Neither by representing theories as opinionated events nor by representing them as opinionated outcomes do we produce something that works as a model of morally uncertain agents, or as a deliberative tool for them. But we must represent them as one or the other for this modelling approach to succeed. The approach has therefore failed.

It is beyond the scope of this article to say how one should model moral uncertainty instead, but I am inclined to speculate that moral uncertainty is just a special case of what in decision theory is called desirabilistic uncertainty, or uncertainty about what to prefer to what.[11] Our mistake has been to treat moral doubt as uncertainty of which moral theory is true rather than as ordinary uncertainty of how desirable different acts are. That is, I'd wager that the uncertainty we experience when we don't know whether to push the man on the trolley tracks is of a piece with our uncertainty when we don't know whether to take the sinfully scrumptious slice of chocolate cake for desert or the healthy apple rather than with our uncertainty of tomorrow's weather. I'd wager it is also of a piece with the uncertainty one feels when one is torn between one's affections for two different possible romantic partners, each of whom arouses our interest and desire. No one seriously proposes that the right way to decide whether to pursue a relationship with Betty or Veronica is to represent each possible romantic partner as a possible description of how the world might stand, romantically, and then try to assign credences to these descriptions. Perhaps it is a mistake to try to do this with morality.

Conscience, after all, is just another human motive, alongside self-interest and affection. It would be surprising if the practical dilemmas posed by the inner turmoil of our conscience had to be modelled radically differently than those posed by the conflict between competing prudential drives or strains of affection. Many ways of modelling this kind of desirabilistic uncertainty that have been proposed, none of which involve representing moral theories as events or possible outcomes. Clues to the modelling of moral uncertainty will be found there. And if I am right about this, then the advice we give to the morally uncertain is the advice we give everyone: if you have sharp credences and preferences, maximize your expected utility; if not, you have a choice of imprecise Bayesian decision theories.

Acknowledgments. My thanks to Richard Bradley and Johanna Thoma for their very helpful comments, and my deepest thanks to Nicholas Baigent, without whose excellent advice I would never have conceived of this project.

Nicolas Côté is a lecturer at the University of Glasgow. His research is mainly in moral and political philosophy, with a focus on the axiomatic foundations of moral theory, the measurement and weighing of moral values, and the interaction between norms of rationality and norms of morality. He is also interested in topics related to liberal tolerance and the ethics of aid and development, and particularly in the conflict between the well-meaning motives that drive aid policy and the imperialist character that those policies often present.

[11]See Lewis (1988, 1996), Broome (1991), Bradley and List (2009), Bradley and Stefànsson (2010), Bradley (2017), and Hill (2013).

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

References

Bradley, Richard., and Christian List. 2009. "Desire-as-Belief Revisited." *Analysis* 69: 31–37.

Bradley, Richard., and Orri. Stefánsson. 2010. "Desire, Belief, and Invariance." *Mind* 125: 691–725.

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

Broome, John. 1991. "Desire, Beliefs, and Expectation." *Mind* 100: 265–67.

Dietrich, Franz, and Brian Jabarian. 2021. "Decision under Normative Uncertainty." *Economics and Philosophy* 38 (3): 372–94.

Enoch, David. 2013. "A Defense of Moral Deference." *Journal of Philosophy*: 229–58.

Gracely, Edward. 2013. "On the Noncomparability of Judgments Made by Different Ethical Theories." *Metaphilosophy* 3: 327–32.

Gustafsson, Johann., and Olle Torpman. 2014. "In Defence of My Favorite Theory." *Pacific Philosophical Quarterly*: 159–74.

Harman, Elizabeth. 2015. "The Irrelevance of Moral Uncertainty." In *Oxford Studies in Metaethics*, vol. 10, edited by Russ Shafer Landau, 53–79. Oxford: Oxford University Press.

Hedden, Brian. 2013. "Does MITE Make Right? Decision-Making under Normative Uncertainty." In *Oxford Studies in Metaethics*, vol. 11, edited by Russ Shafer Landau, 102–28. Oxford: Oxford University Press.

Hicks, Amelia (2021). "Non-ideal prescriptions for the morally uncertain" *Philosophical Studies* 179 (4):1039–1064

Hill, B. 2013. "Confidence in Preferences." *Social Choice and Welfare* 39: 273–302.

Kernohan, Andrew. 2021. "Descriptive Uncertainty and Maximizing Expected Choice-Worthiness." *Ethical Theory and Moral Practice* 24: 197–211.

King, Zoë Johnson. 2022. "Who's Afraid of Normative Externalism?" In *Meaning, Decision, and Norms: Themes From the Work of Allan Gibbard*, edited by Billy Dunaway and David Plunkett. Michigan Publishing Services.

Lewis, David 1980. "A Subjectivist's Guide to Objective Chance." In R. C. Jeffrey(Ed.) *Studies in Inductive Logic and Probability*, vol. II. Berkeley: University of California Press.

Lewis, David. 1988. "Desire-as-Belief." *Mind* 97: 323–32.

Lewis, David. 1996. "Desire-as-Belief II." *Mind* 105: 303–13.

Lockhart, Ted. 2000. *Moral Uncertainty and Its Consequences*. New York: Oxford University Press.

MacAskill, William. 2013. "The Infectiousness of Nihilism." *Ethics* 123: 508–20.

MacAskill, William, and Toby Ord. 2018. "Why Maximize Expected Choice-worthiness?" *Noûs* 54: 1–27.

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

Nissan-Rozen, Ittay. 2015. "*Against Moral Hedging*." Economics and Philosophy 349–69.

Nissan-Rozen, Ittay. 2018. "Is Value under Hypothesis Value?' *Ergo* 5.

Podgorski, Abelard. 2020. "Normative Uncertainty and the Dependence Problem." *Mind* 129: 43–70.

Putnam, Hillary. 2002. *The Collapse of the Fact/Value Dichotomy*. Cambridge, MA: Harvard University Press.

Putnam, Hillary. 2004. *Ethics without Ontology*. Cambridge, MA: Harvard University Press.

Riedener, Stefan. 2020. "An Axiomatic Approach to Axiological Uncertainty. *Philosophical Studies* 177 (2): 483–504.

Riedener, Stefan. 2021. *Uncertain Values: An Axiomatic Approach of Axiological Uncertainty*. Berlin: De Gruyter.

Robinson, Pamela. 2021. "Is Normative Uncertainty Irrelevant If Your Descriptive Uncertainty Depends on It?" *Pacific Philosophical Quarterly* 103 (4): 874–99.

Ross, Jacob. 2006. "Rejecting Ethical Deflationism." *Ethics* 116 (4): 742–68.

Sen, Amartya. 1971. "Choice Functions and Revealed Preference." *Review of Economic Studies* 38: 307–17.

Sepielli, Andrew. 2009. "What to Do When You Don't Know What to Do." *Oxford Studies in Metaethics* 4: 5–28.

Sepielli, Andrew. 2013. "Moral Uncertainty and the Principle of Equity among Moral Theories." *Philosophy and Phenomenological Research* 86 (3): 580–89.

Sepielli, Andrew. 2016. "Moral Uncertainty and Fetishistic Motivation." *Philosophical Studies*: 2951–68.

Steuwer, Bastian. 2021. "Aggregation, Balancing, and Respect for the Claims of Individuals." *Utilitas* 33: 17–34.

Tarnsey, Christian. 2021. "Vive la Différence? Structural Diversity as a Challenge for Metanormative Theories." *Ethics*: 151–82.

Temkin, Larry. 2012. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press.

Tessman, Lisa. 2015. *Moral Failure: On the Impossible Demands of Morality*. Oxford: Oxford University Press.

Voorhoeve, Alex. 2014. "How Should We Aggregate Competing Claims?" *Ethics* 125: 64–87.

Weatherson, Brian. 2014. "Running Risks Morally." *Philosophical Studies* 1: 141–63.

Weatherson, Brian. 2021. *Normative Externalism*. Oxford: Oxford University Press.

Williamson, Timothy. Forthcoming. "Moral Anti-Exceptionalism." In *The Oxford Handbook of Moral Realism*, edited by Paul Bloomfield and David Copp. Oxford: Oxford University Press.

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

Appendices

A Definition of a Boolean algebra and a well-formed theory

The object language $\mathcal{L} = \langle \mathcal{X}, \neg, \wedge, \vee, \bot, T \rangle$ is a Boolean algebra if and only if:

Normality	$a \vee \neg a = T$	$a \wedge \neg a = \bot$
Commutativity	$a \vee b = b \vee a$	$a \wedge b = b \wedge a$
Associativity	$\vee (b \vee c) = (a \vee b) \vee c$	$a \wedge (b \wedge c) = (c \wedge b) \wedge c$
Idempotence	$a \vee a = a$	$a \wedge a = a$
Absorption	$a = a \vee (a \wedge b)$	$a = a \wedge (a \vee b)$
Distributivity	$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$	$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$

We can define the implication relation $\vDash$ on $\mathcal{X}$ by

$$a \vDash b \Leftrightarrow a \vee b = b \Leftrightarrow a \wedge b = a.$$

A well-formed theory A has the following properties.

$$\wedge_{a \in \mathcal{A}} \nvDash \bot \tag{1}$$

$$\wedge_{a \in \mathcal{A}} \vDash b \Rightarrow b \in \mathcal{A} \tag{2}$$

$$\forall S \subseteq A, x \in S \Rightarrow \text{either } \Diamond(x, S) \in \mathcal{A} \text{ or } \neg\Diamond(x, S) \in \mathcal{A} \tag{3}$$

$$\forall S \subseteq A, \exists x \in S : \Diamond(x, S) \in \mathcal{A} \tag{4}$$

B Proof that $\mathcal{F}$ is the smallest σ-algebra to include all preorders.

Let Dave be an arbitrary individual and $\mathcal{F}_{\mathrm{d}}$ be the set of all relations on A that are either reflexive or irreflexive. It is easily shown that $\mathcal{F}_d$ is closed under complement and countable union. Suppose R_d and T_d are both reflexive, then $R_d \setminus \mathrm{T_d}$ is irreflexive and therefore an element of $\mathcal{F}_d$. Likewise if they are both irreflexive. If R_d is reflexive and T_d is irreflexive, then $R_d \setminus \mathrm{T_d}$ is reflexive and $T_d \setminus \mathrm{R_d}$ is irreflexive. Likewise, the union of any countable number of relations that includes at least one reflexive relation is reflexive, while the union of any countable number of irreflexive relations is irreflexive. Closure under countable union implies closure under countabile intersection, so $\mathcal{F}_{\mathrm{d}}$ is a σ-algebra.

$\mathcal{F}_{\mathrm{d}}$ is also the smallest σ-algebra to include all Dave-indexed preorders. Suppose $\mathcal{F}$ is a σ-algebra that includes all Dave-indexed preorders, and let $(a,b) \in A \times A : a \neq b$ be an arbitrary element of $A \times A$. The set of all Dave-indexed preorders on A includes $R_{\mathrm{d}} = \{(x,y) \in A \times A | (x,y) = (a,b) \text{ or } x = y\}$ and $T_{\mathrm{d}} = \{(x,x) \in A \times A\}$ (these are both reflexive and vacuously transitive). Therefore $R_{\mathrm{d}}, T_{\mathrm{d}} \in \mathcal{F}$, and since $\mathcal{F}$ is a σ-algebra, $R_d \setminus \mathrm{T_d} = \{(\mathrm{a}, \mathrm{b})\} \in \mathcal{F}$. Since (a,b) was arbitrary, it follows that for all $(x,y) \in A \times A : x \neq y, \{(x,y)\} \in \mathcal{F}$. Note that $\{(x,y)\}$ in this case is irreflexive. Since $\mathcal{F}$ is closed under countable union, it follows that any nonempty irreflexive relation is an element of $\mathcal{F}$, and obviously the

https://doi.org/10.1017/can.2023.18 Published online by Cambridge University Press

empty set is an element of $\mathcal{F}$ since $\mathcal{F}$ is a σ-algebra, so all irreflexive relations on A are elements of $\mathcal{F}$. Moreover, any relation which is the union of an irreflexive relation and a preorder is an element of $\mathcal{F}$. Note that any reflexive relation is the union of a preorder and an irreflexive relation: let R be an arbitrary reflexive relation on A, and define $R_1, R_2 \subseteq R$ such that $R_1 = \{(x,x) \in A \times A\}, R_2 = \{(x,y) \in A \times A | (x,y) \in R \text{ and } x \neq y\}$. R_1 is vacuously a preorder and R_2 is an irreflexive relation. So $\mathcal{F}$ includes at a minimum all relations which are either reflexive or irreflexive, i.e., $\mathcal{F}_\mathrm{d} \subseteq \mathcal{F}$. Therefore $\mathcal{F}_\mathrm{d}$ is the smallest σ-algebra to include all Dave-indexed preorders.

But $\mathcal{F}_\mathrm{d}$ is identical with the set of all relations on A that are either reflexive or irreflexive. Since Dave was an arbitrary individual, it follows that for any individual i, the smallest σ-algebra to include every i-indexed preorder is the family $\mathcal{F}$ of all relations that are either reflexive or irreflexive.

Cite this article: Côté, N. 2023. There Is No Such Thing as Expected Moral Choice-Worthiness. *Canadian Journal of Philosophy* 53: 1–20, doi:10.1017/can.2023.18