A Computable von Neumann-Morgenstern Representation Theorem

Josiah Lopez-Wild

1 Introduction

Decision theory seeks to define rational choice behavior. Given a collection of acts available to some decision maker, decision theorists commonly identify the "rational" act as the act which maximizes expected utility (where the expectation is taken with respect to some probability measure). As a simple case, suppose P and Q are gambles—that is, probability distributions over some common outcome space X. Suppose $u : X \to \mathbb{R}$ is a *utility function*, a function that assigns a real number (representing its "value" to the decision maker) to each possible outcome. A higher utility corresponds to a more desirable outcome. Then the (standard) decision theorist claims that a rational agent prefers P to Q just in case

$$\mathbb{E}_P[u] \geq \mathbb{E}_Q[u].$$

In any standard reference text on expected utility theory one will find *representation theorems* (for example, [27], [13], [35]). These theorems link expected utility maximization to a qualitative description of an agent's choice behavior. Typically an agent's choice behavior is captured by a *preference relation* $\preceq$ on the set of decisions they face (in our above example this is the set of gambles, but preferences might instead be defined on acts which have no intrinsic probabilities). We say that $P \preceq Q$ if and only if the agent deems Q to be at least as desirable as P. We then prove something of the form: the preference relation satisfies a given set of axioms if and only if there exists a utility function (and, in some cases, a probability measure) such that the agent prefers gambles with greater expected utility ([32], [42]). The most prominent instances of such representation theorems are due to von Neumann and Morgenstern ([38]), Anscombe and Aumann ([1]), and Savage ([35]).[1]

[1] One should also mention the work of Ramsey ([34]) as a predecessor to these theorems.

This paper has two primary aims. First, we are interested in These theorems are generally taken to be foundational to decision theory in some way. The kind of foundational role the theorems play depends on how one interprets decision theory as a whole. In §2 we will consider two dimensions: the normative/descriptive dimension and the mentalist/behaviorist dimension. These two dimensions are orthogonal and one could in principle adopt any pair of views. Representation theorems play different roles under each pair of views, as we will see.

2 The Role of Representation Theorems in Decision Theory

In this section we will distinguish a number of ways that one can approach the enterprise of decision theory and the resulting roles of representation theorems.[2] These distinctions are widespread in the decision theory literature ([32], [10], [24], [42], [26]). I will not argue for or against the philosophical fruitfulness of the distinctions; the goal of this section is to provide a common language to understand the motivation for representation theorems and by extension the current project.

Let's begin by stating the von Neumann-Morgenstern representation theorem. This will allow us to be more concrete in the ensuing discussion. Suppose we have a set X of outcomes and a set $\mathscr{P}(X)$ of probability distributions on X. The elements of $\mathscr{P}(X)$ are "gambles" or "lotteries" on X. Suppose also there is a relation $\preceq$ on $\mathscr{P}(X)$; this is the agent's "preference relation", and we write $P \preceq Q$ to mean that the agent thinks Q is at least as good as P. We can write $P \prec Q$ to mean that the agent thinks Q is strictly better than P, and $P \sim Q$ to mean that the agent thinks P and Q are equally desirable.

von Neumann and Morgenstern assume the following axioms are true of $\preceq$:

(R1) $\preceq$ is reflexive, transitive and totally connected.

(R2) If $P \preceq Q$ and $\alpha \in [0, 1]$ then $\alpha P + (1-\alpha)R \preceq \alpha Q + (1-\alpha)R$ for all $R \in \mathscr{P}$.

(R3) If $P \preceq Q \preceq R$ then there exist $\alpha, \beta \in (0, 1)$ such that $\alpha P + (1-\alpha)R \preceq Q \preceq \beta P + (1-\beta)R$.

In other words, (R1) states that $P \preceq P$ for all gambles P, if $P \preceq Q$ and $Q \preceq R$ then $P \preceq R$, and for any two gambles P, Q either $P \preceq Q$ or $Q \preceq R$. (R2) supposes that $P \preceq Q$ and that we flip a coin with bias α toward heads. One gamble, $\alpha P + (1-\alpha)R$, gives us gamble P if the coin is heads and R if it is tails, whereas $\alpha Q + (1-\alpha)R$ gives us gamble Q on heads and R on tails. Since the agent thinks Q is at least as good as P and the odds of getting R are the same in both compound gambles, the agent thinks the compound gamble involving Q is at least as good as that involving P.

[2]See also Bermúdez ([3], especially Chapter 2) for a discussion of the mentalist/behaviorist distinction and its effects on the role of representation theorems.

This is a sort of independence axiom; the only relevant difference between the two gambles is P and Q, and R has no effect on their desirability. Finally, (R3) says that if $P \preceq Q \preceq R$ then we can make compound gambles of P and R that either (i) make P so likely that the gamble is no *better* than Q, or (ii) make R so likely that the gamble is no *worse* than R. This is an "Archimedean" axiom; it says that no gambles are infinitely better or worse than any others.

Assuming these axioms von Neumann and Morgenstern prove that an agent's preferences satisfy axioms (R1)–(R3) if and only if there is a function $u : X \to \mathbb{R}$ such that

$$P \preceq Q \iff \mathbb{E}_P[u] \leq \mathbb{E}_Q[u].$$

That is, if an agent's preferences satisfy those axioms then there must be some utility function that the agent maximizes in expectation.

To understand why this theorem (and similar theorems which came after it) are so important in decision theory, we need to determine how we understand decision theory itself. In philosophy it is common to treat expected utility theory as a *normative* theory, i.e., one that describes how rational agents *should* behave. We think of expected utility maximization as constitutive of rational behavior—one ought to maximize expected utility. If we are interested in a normative theory then we must ask: where does this normative force come from? More specifically one might wonder why rationality requires the *maximization* of *expected utility* and not some other quantity, or some other operation besides maximization. Some philosophers have argued that expected utility maximization is simply "rational bedrock" ([7]) which requires no further explanation. For example Lewis says that decision theory "is not esoteric science, however unfamiliar it may seem to an outsider. Rather, it is a systematic exposition of the consequences of certain well-chosen platitudes about belief, desire, preference, and choice" ([28, 338]). Others have argued that one should maximize something *other* than expected utility; for example, Buchak ([9]) has developed an alternative to expected utility that includes an agent's attitude toward *risk*. An agent who is risk-averse, for example, might put more weight on unlikely but exceptionally bad outcomes than would an agent who simply maximizes expected utility, resulting in different recommendations for the "rational" act.[3]

One might instead try to justify, rather than merely stipulate, the claim that rational acts are exactly the acts which maximize expected utility. One prominent justification strategy is to appeal to a representation theorem. On this strategy one takes the axioms of the representation theorem to *ground* the normative force of expected utility theory by arguing that the axioms themselves are intuitively rational ([35, 97], [10], [32, 414]). For example, von Neumann and Morgenstern's axioms stipulate requirements such as transitivity (R1): if an agent prefers P to Q, and prefers Q to

[3] See [8] for a comprehensive overview of other alternatives to expected utility theory.

R, then they prefer P to R. This axiom strikes many as intuitively rational. Since any agent whose preferences satisfy such axioms is thereby representable as maximizing expected utility, it follows that, to the extent that one finds the axioms to be intuitively rational, maximizing expected utility is itself rational.

By contrast we may be interested in using decision theory in a descriptive capacity. Decision theory plays a descriptive role when we treat it as a scientific theory that may or may not be true of certain parts of the world (in this case, agents and their decisions). This is view of decision theory is the norm in economics, psychology, cognitive science, and elsewhere. In economics this approach is well-known in the form of the "expected utility hypothesis" (see [16] for a classic discussion). This hypothesis states that agents maximize expected utility when faced with decisions under uncertainty. As Savage and Friedman put it, expected utility theory

> asserts that individuals behave as if they calculated and compared expected utility and as if they knew the odds… the validity of this assertion does not depend on whether individuals know the precise odds, much less on whether they say that they calculate and compare expected utilities or think that they do, or whether psychologists can uncover any evidence whether they do, but solely on whether it reveals sufficiently accurate predictions about the class of decisions with which the hypothesis deals ([15, 282]).

In other words, it is an *empirical hypothesis*; we are interested in how well expected utility maximization predicts the behavior of real agents. A large body of work in economics is devoted to determining whether (i) agents do in fact maximize expected utility,[4] and (ii) what sorts of utility functions best describe real agents.[5]

The Savage and Friedman quote is explicitly descriptive—their interest is in whether the expected utility representation "yields sufficiently accurate predictions" of an agent's behavior. In this case we are interested in testing the validity of the expected utility hypothesis. A representation theorem allows us to test the hypothesis by checking whether an agent's preferences satisfy the axioms of the theorem. If the preferences do satisfy the axioms, then, since they are sufficient for the representation, the agent *does* act as if they maximize expected utility; if the preferences do not satisfy all the axioms, then, since they are necessary, the agent does *not* act as if they maximize expected utility.

Note also that Friedman and Savage claim that the validity of the expected utility hypothesis does not depend on the agent having explicit representations of probabilities or utilities, nor on them calculating expected utility before acting. In other words these authors take a *behaviorist* stance

[4] See especially the field of behavioral economics, including the literature on prospect theory ([25]).

[5] See [2], [33].

toward decision theory;[6] the preference relation is considered as the primitive entity, and the agent acts "as if" they maximize the corresponding utility function. The utility function need not be a psychologically real entity, and the agent is not assumed to introspect on the utility of outcomes or calculate the expected utility of gambles. Some philosophers take this stance too: Joyce writes that

> No sensible person should ever propose expected utility maximization as a *decision procedure*. nor should he suggest that rational agents must have the maximization of utility as their *goal*... The expected utility hypothesis is a theory of "right-making characteristics" rather than a guide to rational deliberation. It in no way requires an agent *consciously* to assign probabilities to states of the world or utilities to outcomes, or to actually calculate anything. The decision maker does not need to have a concept of utility at all, and she certainly does not have to see herself as an expected utility maximizer. The demand is merely that her desires and beliefs, however arrived at, should be *compatible* with the expected utility hypothesis in the sense that it should be possible for a third party who knows her preference ranking to represent it in the way described ([24, 80]).

A representation theorem cashes out the precise sense in which an agent has preferences which a third party could use to represent their behavior as rational.

This view stands in contrast to a *mentalist* conception of decision theory, on which probabilities and utilities are psychologically real entities. These probabilities and utilities need not be introspectible by the agent, however. If they are not introspectible then they must be discovered indirectly, perhaps via a representation theorem or other measurement procedure.[7] If they are introspectible then expected utility theory prescribes that the agent actually calculate the expected utility of their acts, and the fact that an agent prefers gamble P to Q can be explained by the fact that P has greater expected utility (with respect to the agent's personal utility and probability functions). For example, Lewis takes as primitive the agent's probabilities and utilities for his discussion, and appears to think of them as mental entities when he says "it seems most unlikely that any real person could store and process anything so rich in information as the C [probability] and V [utility] functions envisaged... But it is plausible that someone who really did have these functions to guide him would not be so very different from us in his conduct" ([29, 7]). So for Lewis an agent's probabilities and utilities are meant to *guide* the agent's action, and so are presumably introspectible. Moreover Lewis' project, like many philosophers', is normative rather than descrip-

[6] Note that I do *not* mean "behaviorist" to mean psychological behaviorism, but rather to mean anti-realism about the derived utility or probability functions, as I explain below.

[7] Khan ([26]) calls this position "weak mentalism", as distinguished from the "strong mentalist" position on which probabilities and utilities are introspectible. Ramsey ([34]) may have been an early proponent of this sort of weak mentalist view.

tive. For a mentalist a representation theorem can play various roles. A representation theorem can provide justification for the claim that one *should* maximize expected utility. Alternatively such a theorem provides a simpler set of rules that an agent's preferences must obey if they maximize expected utility, so can serve as an additional check on an agent's behavior.

A prominent contemporary strand of mentalist positions can be found in Bayesian cognitive science and psychology (see e.g. [31], [30], [17], [12]). These projects are aimed at showing that human behavior, especially cognitive processes such as planning, problem solving, vision, motor control, etc., can be fruitfully modeled as some kind of (bounded) expected utility maximization. In this case one assumes that there is some explicit representation of probability and utility that the agent (or some cognitive subsystem of the agent) has access to in order to calculate the optimal act. These projects are inherently *descriptive*, as their goal is to provide empirical models of actual human cognition.

Thus the two dimensions we have discussed (normative/descriptive and behaviorist/mentalist) are orthogonal and one could in principle hold any pair of views. Moreover representation theorems have roles to play regardless of how one approaches decision theory. It is clear why representation theorems have formed part of the foundation of decision theory since von Neumann and Morgenstern.

3 Computability in Decision Theory

In the previous section we saw that on all major interpretations of decision theory, representation theorems provide some kind of foundation. That said, many have been critical of the assumptions and scope of these theorems. Some have argued that certain axioms are unrealistically strong and so the theorems do not apply to real agents. For example, the von Neumann-Morgenstern theorem assumes that an agent's preferences are *connected*, that is, that for any two gambles P, Q, either $P \preceq Q$ or $Q \prec P$. Some authors have argued that this is unrealistic, either because it is too cognitively demanding for an agent to maintain preferences over *all possible* gambles, or because some alternatives will simply be incommensurable. Alternative representation theorems that discard various assumptions have been proved ([36]).

One intuition underlying these criticisms is that being fully rational—having rational preferences or maximizing expected utility—is somehow "hard", too hard for real finite agents like ourselves. This point has been raise by philosophers ([19], [41], [11]) and scientists alike ([25]), and is perhaps best represented in the traditions of bounded rationality ([37]) and modern Bayesian cognitive science and psychology. These programs take standard decision theory as an ideal to be *approximated*. The precise details of this approximation vary. One might designate some threshold expected utility which is "good enough" (an approach called "satisficing", due to Simon).

Instead one might instead fix some set of "computational resources" that an agent must spend when computing expected utility. In this latter case one determines how well an agent can perform given the prescribed constraints. Again these resources can vary depending on the project. It is common to assume a computational model such as a Turing machine; one then stipulates that the Turing machine can only run in a bounded number of steps (a "time" constraint) or that it can only use a certain number of cells on its work tape (a "space" constraint), or that its tape is of fixed finite length (a "finite-state automoton").

Proceeding in this manner brings decision theory into contact with computational complexity theory, the mathematical study of problems which are solvable via Turing machine in a limited amount of time or with a limited amount of space. This is distinguished from classical computability theory, which studies the problems that are solvable given arbitrary (but still finite) time and space. Complexity theory is valuable in the design and analysis of actual algorithms we use on a day to day basis—including, in this case, our own cognitive processes. We require efficient algorithms that can be performed on finite hardware such as processors or neurons and that return results in a timely manner. Computability theory's main philosophical interest lies in its ability to show that certain problems are unsolvable *even if* we had arbitrarily powerful computational machines. Complexity theory wins its usefulness in real-world application at the expense of immense difficulty—many foundational questions remain open to this day. Computability theory, by contrast, is far better understood in part because it idealizes away from time and space considerations.

Despite the fact that computational complexity plays an increasingly important role in empirical applications of decision theory, this role has been little discussed by philosophers (though see [22], [23]). This may be due to a number of factors. First, computational complexity in general does not receive much attention from philosophers. Second, philosophical interest in decision theory is quite general: philosophers tend to be interested in providing ideal norms that rational agents should aspire to, regardless of their contingent limitations ([11]).

Given these considerations I argue that *computability* theory, rather than computational complexity, is a principled middle ground that is capable of describing more realistic agents—thus responding to concerns that expected utility maximization is "too hard" for real agents—while also being sufficiently general that one can draw philosophical morals from the results.[8] To my knowledge there has not appeared any systematic discussion of computability in decision theory in the philosophical literature. My goal is to initiate this study. I will begin with the foundations: a representation theorem. The remainder of this section explains why a computable representation theorem is of interest for a decision theory of realistic agents.

First, most real systems we would be willing to call "agents" in the decision-theoretic sense—

[8] I also believe that computational complexity theory contains overlooked philosophical morals, but that argument must wait for another paper.

humans, nonhuman animals, possibly some algorithms—are most likely *computable* agents. If the above programs in Bayesian cognitive science are correct then we can make the stronger claim that humans (and hence nonhuman animals) are bounded by *efficient* computability, not mere computability *simpliciter*. The same is obviously true of any reasoning or inference algorithms in current use, e.g. machine learning models. So assuming that our intended agent's reasoning capacities are bounded by Turing computability is a realistic and quite permissive assumption.

What would it mean for an agent to be computable? Since we are interested in decision theory, this would mean that the agent's probability and utility functions must be computable functions. We defer a precise definition for later in the paper, but the intuitive idea is that a probability function is computable if there is an algorithm which takes a coded description of some event and returns (a code for) the probability of that event; likewise a utility function is computable if there is an algorithm that takes codes for gambles or events and returns a code for the utility of that gamble/event.

The classical von Neumann-Morgenstern theorem answers the question "When can an agent be represented as maximizing expected utility?" If we are interested in computable agents we might similarly ask "When can an agent be represented as maximizing a *computable* utility function, in expectation?" In other words, when can we safely represent an agent as computable? What are the most general conditions under which this is possible? This question will be answered by a computable version of the von Neumann-Morgenstern representation theorem.

But we can give a distinct motivation, which is surprisingly resolved by the same theorem. Suppose you are a scientist who wishes to use expected utility theory in a descriptive capacity. That is, you wish to use the calculation of expected utility to predict agents' future choice behavior.[9] A representation theorem such as von Neumann and Morgenstern's tells us that this can be done if an agent's preferences satisfy the axioms. In that case, the theorem says, there is a utility function that you can use to calculate what an agent values. If one assumes that the outcome space X and the space of probabilities $\mathscr{P}(X)$ is finite, then one can also write an explicit definition of a suitable utility function.

While in philosophy decision theorists often work in finite spaces in order to make their arguments more transparent, it is common practice in economics to define utility functions on infinite spaces—for example, one often defines utilities on intervals of the real line, or real-valued random variables representing uncertain commodities. Moreover when X is infinite it is often desirable to consider the infinite set $\mathscr{P}(X)$ of all gambles on X. Examples of gambles found in textbooks often employ so-called "simple" probability measures with finite support that assign

[9]Friedman and Savage again: "Given a utility function obtained [via the vNM representation theorem], it is possible, if the [expected utility] hypothesis is correct, to compute the utility attached to (that is, the expected utility of) any set or sets of possible incomes and associated probabilities and thereby to predict which of a number of such sets will be chosen" ([15, 292]).

rational numbers to each point in their support. One might think we could therefore ignore other probability measures and work with the simpler set of simple measures. This would, however, rule out very natural probability distributions for our gambles. The expected utility of lotteries with normal distributions, which are commonly used in practice, would not be defined in that case. So there is good reason to work with the entire set of probability measures on X.

But in this general case the classical von Neumann-Morgenstern theorem tells us only that *there is* a suitable utility function, given an agent's preferences over given probability distributions. It does not, in general, explicitly tell us *how to find* this function. Put slightly differently, we are not given an algorithm which, given the agent's preferences, computes a utility function for that agent. This is a problem for our scientist's project. To predict an agent's behavior (according to the theory) we must calculate expected utility, and to calculate expected utility we must actually possess a utility function. In other words, the von Neumann-Morgenstern utility function must be *computable from* the agent's preference relation.[10] Conversely, if the utility function were not computable from the preference relation, then there may be instances where we cannot actually determine which utility function represents the agent. This inability would render the descriptive project impossible in some cases; we would not be in a position to determine whether the agent acts according to the principle of expected utility maximization, and would not be able to predict their future behavior on this basis. We are therefore interested in the question: given access to an agent's preference behavior, can we compute a utility function that represents that behavior?

Surprisingly this second motivation will be answered by one and the same theorem. I provide highly general sufficient conditions for the computability of a von Neumann-Morgenstern utility function from a preference relation. To do so I use tools from the field of computable analysis. In §5 I introduce the notion of a *computable continuous preference relation* in Definition 4. §6 builds up to the statement of the main theorem, Theorem 2, which proves that a computable utility function exists if the preference relation $\preceq$ is computable continuous. The proof of this theorem shows more generally that the agent's preference relation, if treated as an oracle, is sufficient to compute a von Neumann-Morgenstern utility. Most definitions and essential ideas are found in the main body of the text, but all proofs have been moved to an appendix.

[10]One might argue that mere computability is not sufficient for our economist's descriptive program. After all, a working economist needs not only an algorithm that produces a utility function, but also an algorithm that does so *efficiently*. So there is a further need to consider the computational complexity of any algorithm that produces a utility function. This is certainly true, and promises interesting work. We can view the present project as providing general conditions under which it is *possible* to ask for efficient algorithms for producing a utility function. Finding those algorithms must be left for future work.

4 Computability and Order on Countable Spaces

We begin with a few mathematical preliminaries. We say that a binary relation $\preceq$ is a *preorder* if it is reflexive and transitive. We say that $\preceq$ is a *total preorder* if it is a preorder and is connected. We call a set X together with a binary relation $\preceq$ an *ordered set*. From $\preceq$ we define

$$P \prec Q \iff P \preceq Q \wedge Q \not\preceq P,$$
$$P \sim Q \iff P \preceq Q \wedge Q \preceq P.$$

In what follows we will represent an agent's preferences as a total preorder. Intuitively this means that an agent has a preference between any two options—there are no gambles P, Q that are incommensurable—but ties are allowed.

An ordered set $(X, \preceq)$ is *dense* if for any $x, y \in X$ such that $x \prec y$, there is $z \in X$ such that $x \prec z \prec y$. Further, $(X, \preceq)$ is *unbounded* if for any $x \in X$ there are $y, z \in X$ such that $z \prec x \prec y$. For an ordered set $(X, \prec)$ we write $(\leftarrow, x)$ and $(x, \rightarrow)$ to denote the sets $\{y \in X \mid y \prec x\}$ and $\{y \in X \mid x \prec y\}$, respectively. Given two sets $(X, \prec_X), (Y, \prec_Y)$ equipped with binary relations, a function $h : X \to Y$ is an *order isomorphism* if h is a bijection and for all $x_1, x_2 \in X$, $x_1 \prec_X x_2 \iff h(x_1) \prec_Y h(x_2)$.

For computability we use standard notation (see for example Soare). A function $f : \omega \to \omega$ is *computable* if there is an index $e \in \omega$ such that $\varphi_e(n) = m \iff f(n) = m$ for all $n \in \omega$. Given a countable set X together with an enumeration of its elements $\alpha : \omega \to X$ and a relation $R \subseteq X \times X$, we say that R is a *computable relation* if $\{(i, j) \in \omega \times \omega \mid (\alpha(i), \alpha(j)) \in R\}$ is computable. When working with computable relations we may equivocate between the element $x \in X$ and its index $i \in \omega$ such that $\alpha(i) = x$, and frequently write x_i. A pair $(X, \prec)$ where X is a countable set and $\prec$ is a computable binary relation is called a *computably ordered set.*

5 Computable Analysis

In §3 I argued that it is commonplace in fields such as economics to define utility functions over infinite outcome spaces.[11] Thus I will choose to work in a highly general setting to accurately model that practice. Specifically the outcome set X is usually uncountable in practice (e.g. whenever a utility is defined on the real numbers). The traditional definition of Turing computability is only well-defined on countable sets, so it no longer applies. Thus we must introduce ideas from computable analysis to extend computability notions to these larger spaces. The most popular foundation for computable analysis is Weihrauch's Type-II Theory of Effectivity (TTE), as presented

[11] Indeed, standard textbooks such as [13], [27] spend significant time showing how this is done.

in [40], [6].[12] The most common setting for work in computable analysis is a *computable Polish space.*

Definition 1 (Computable Polish space)**.** A *computable Polish space* is a triple $(X, \mathscr{D}, d)$ such that

1. X is a separable complete metric space with metric d;
2. $\mathscr{D}$ is a countable dense subset of X with an enumeration $\{s_n\}_{n\in\omega}$;
3. for all $s_n, s_m \in \mathscr{D}$ we have that $d(s_n, s_m)$ is a computable real, uniformly in n, m.

Some natural examples of computable Polish spaces include $\mathbb{R}, 2^\omega, \omega^\omega$, and $\mathbb{Q}$. Importantly, if X is a computable Polish space then one can show that the space $\mathscr{P}(X)$ of (Borel) probability measures on X is a computable Polish space ([21]). In this case the countable dense set $\mathscr{D}_{\mathscr{P}(X)}$ is the set of all *simple* measures on $\mathscr{D}$, i.e., all probability measures μ of the form

$$\mu = \sum_{n\in I} q_n \delta_{s_n}$$

where $I \subseteq \omega$ is finite and δ_{s_n} is the probability measure that assigns measure 1 to the set $\{s_n\}$. For the metric of $\mathscr{P}(X)$ we use the *Prokhorov metric* given by

$$d_P(\mu, \nu) = \inf\{\epsilon > 0 \mid \nu(B) \leq \mu(B^\epsilon) + \epsilon \ \wedge \ \mu(B) \leq \nu(B^\epsilon) + \epsilon \text{ for all Borel sets } B\}$$

for all $\mu, \nu \in \mathscr{P}(X)$ ([4], [5]). Thus $(\mathscr{P}(X), \mathscr{D}_{\mathscr{P}(X)}, d_P)$ is a computable Polish space whenever X is.

Since the members of $\mathscr{D}_{\mathscr{P}(X)}$ are *finite* mixtures we may represent them via n-tuples $(q_0, s_0, q_1, s_2, \ldots q_n, s_n)$ of rational weights $q_n \in \mathbb{Q}$ and points s_n from $\mathscr{D}$. Thus if $\langle \cdot, \cdot \rangle : \omega \times \omega \to \omega$ is a computable pairing function we may take the enumeration $\{D_n\}_{n\in\omega}$ of $\mathscr{D}_{\mathscr{P}(X)}$ to be given by

$$D_i = \sum_{i_n\in I} q_{i_n} \delta_{s_{i_n}} \iff \langle q_{i_1}, s_{i_1}, \ldots q_{i_n}, s_{i_n} \rangle = i$$

with $I \subseteq \omega$ finite and $\langle \cdot, \cdot, \ldots, \cdot \rangle$ given by $\langle \langle \langle \cdot, \cdot \rangle, \cdot \rangle, \ldots, \cdot \rangle$. In this way we may take such n-tuples as the indices on which a Turing machine operates.

Given a computable Polish space X, there is a collection of particularly nicely behaved sets known as *c.e. open sets.*

[12] Other approaches to the foundations of computable analysis exist, some of which are equivalent to TTE and some of which are not. These include domain theory (cite) and a "point-free" approach (). See also (Rute) for a nice comparison of the various approaches.

Definition 2 (C.e. open set)**.** Let X be a computable Polish space. A set $A \subseteq X$ is *c.e. open* if there is a c.e. set $I \subseteq \omega \times \mathbb{Q}$ such that

$$A = \bigcup_{(i,q)\in I} B(s_i, q)$$

where $s_i \in \mathscr{D}$ and $B(s_i, q)$ is the basic open ball centered at s_i with radius q.

Clearly all c.e. open sets are open in X.[13] We think of c.e. open sets as the "effectively specifiable" open sets. Using them we can define a notion of computability of functions on computable Polish spaces. It is a fundamental theorem of computable analysis that all computable functions are continuous ([40]). If $f : X \to Y$ is a function between topological spaces X, Y, then we say that f is *continuous* if for all open $V \subseteq Y$, $f^{-1}(V) \subseteq X$ is open in X. In words, inverse images of open sets are open. We can give an analogous definition of computable functions.

Definition 3. Let X, Y be computable Polish spaces. We say that a function $f : X \to Y$ is *computable continuous* if inverse images of c.e. opens are uniformly c.e. open.

We motivated the idea of a computable function with the picture of an algorithm that takes codes for its arguments and outputs codes for its values. Standard references on computable analysis (e.g. [40], [6]) make this definition precise and prove that it is equivalent to Definition 3.

Computable continuous functions have many nice properties which we will use in subsequent proofs. One particularly important property is that a *real-valued* function $f : X \to \mathbb{R}$ with X a computable Polish space is computable continuous if and only if for all $q \in \mathbb{Q}$ we have that $f^{-1}(-\infty, q)$ and $f^{-1}(q, \infty)$ are c.e. open sets uniformly in q.[14]

We now introduce our notion of a computable continuous *relation* on computable Polish spaces. We say that a relation $\preceq$ on a topological space X is *continuous* if the sets $(\leftarrow, x), (x, \rightarrow)$ are open in X for all $x \in X$.

Definition 4. Given a computable Polish space X and a binary relation $\preceq$ on X, we say that $\preceq$ is *computable continuous* if

1. $\preceq$ is continuous, and

2. for all $d \in \mathscr{D}$, $(\leftarrow, d)$ and $(d, \rightarrow)$ are c.e. open uniformly in x.

Recall that $\mathscr{D}$ is the set of "simple" points of X. Intuitively, then, a relation is computable continuous if whenever given a simple point $d \in \mathscr{D}$, there is an algorithm that enumerates everything

[13] For those familiar with descriptive set theory, c.e. open sets are the analogues of (lightface) Σ^0_1 subsets of ω^ω.
[14] This is because of the simple fact that the sets $(-\infty, q), (q, \infty)$ generate the c.e. open subsets of $\mathbb{R}$.

above d and an algorithm that enumerates everything below d. For our intended use case this means that an agent is capable of listing, in a procedural manner, everything they strictly prefer to some simple gamble and everything they strictly prefer that simple gamble to.

When $\preceq, \prec, \sim$ are restricted to the countable dense set $\mathscr{D}$ we will write $\preceq_{\mathscr{D}}, \prec_{\mathscr{D}}, \sim_{\mathscr{D}}$, respectively. If $\preceq$ is computable continuous then it follows quickly that for all $d \in \mathscr{D}$ the sets $(\leftarrow, d)$ and $(d, \rightarrow)$ are c.e., and $[d] = \{x \in X \mid d \sim x\}$ is co-c.e.

6 Computable Expected Utility

We now turn to our main topic, the von Neumann-Morgenstern representation theorem. As we saw in §2, the vNM theorem states that under certain conditions the existence of a total preorder $\preceq$ on a set of probability measures $\mathscr{P}(X)$ on some space X is equivalent to the existence of a utility function $u : X \to \mathbb{R}$ such that for all $P, Q \in \mathscr{P}(X)$,

$$P \preceq Q \iff \mathbb{E}_P[u] \leq \mathbb{E}_Q[u]. \tag{1}$$

In words, an agent prefers one gamble to another just in case the expected utility of the preferred gamble is greater. To see how this theorem works we'll first introduce an abstract structure called a "mixture space", originally due to Herstein and Milnor [20].

Definition 5. A *mixture space* is a set $\mathscr{P}$ such that for all $\alpha \in [0, 1]$ and $P, Q \in \mathscr{P}$, there is an element $\alpha P + (1 - \alpha)Q \in \mathscr{P}$, which furthermore satisfies

(M1) $1P + 0Q = P$;

(M2) $\alpha P + (1 - \alpha)Q = (1 - \alpha)Q + \alpha P$;

(M3) $\alpha[\beta P + (1 - \alpha)Q] + (1 - \alpha)Q = \alpha\beta P + (1 - \alpha)\beta Q$.

If X is a computable Polish space then $\mathscr{P}(X)$, the space of Borel probability measures on X, is a mixture space, as the reader can verify. Mixture spaces allow one to prove a very general precursor of the vNM theorem, which we call the "Mixture Space Theorem". This result is a halfway step to the full vNM theorem; one can see that the function v is analogous to the expectation of a utility function u.

Proposition 2 (Mixture Space Theorem)**.** *Suppose $\mathscr{P}$ is a mixture space and $\preceq$ is a binary relation on $\mathscr{P}$. Then*

(R1) $\preceq$ is a total preorder;

(R2) if $P \preceq Q$ and $\alpha \in [0,1]$ then $\alpha P + (1-\alpha)R \preceq \alpha Q + (1-\alpha)R$ for all $R \in \mathscr{P}$; and

(R3) if $P \preceq Q \preceq R$ then there exist $\alpha, \beta \in (0,1)$ such that $\alpha P + (1-\alpha)R \preceq Q \preceq \beta P + (1-\beta)R$,

if and only if there exists a function $v : \mathscr{P} \to \mathbb{R}$ such that

$$P \preceq Q \iff v(P) \leq v(Q), \text{ and} \tag{2}$$

$$v(\alpha P + (1-\alpha)Q) = \alpha v(P) + (1-\alpha)v(Q). \tag{3}$$

Moreover, if v represents $\preceq$ in the sense of (4) and (5), v' is another representation if and only if $v' = \alpha v + \beta$ for constants $\alpha > 0$ and β.

Proof. See Appendix. □

Our first theorem shows that one can derive a computable analogue to Proposition 2. In particular, if we assume that the preference relation is computable continuous then we can show that the resulting function v is itself computable continuous.

Theorem 1 (Computable Mixture Space Theorem)**.** *Let X be a computable Polish space, and let $\mathscr{P}(X)$ be the space of Borel probability measures on X. Assume*

(R1) $\preceq$ is a computable continuous total preorder on $\mathscr{P}(X)$,*

and assume (R2), (R3) hold. Then there exists a computable continuous function $v : \mathscr{P} \to \mathbb{R}$ that satisfies (4) and (5).

Proof. See Appendix. □

An important consequence of the classical vNM representation theorem is that the utility function is not unique, but is unique up to positive linear transformation. That is, if u is a utility function that represents an agent's preferences, then so is $u' = au + b$ for real numbers $a > 0, b$. This uniqueness already appears as a consequence of the mixture space theorem. We can prove a computable analogue, namely that our computable continuous function v is unique up to *computable* positive linear transformation.

Corollary 1. *The function $u : X \to \mathbb{R}$ constructed in Theorem 1 is unique up to computable positive linear transformation; that is, $u^* : X \to \mathbb{R}$ satisfies (4) and (5) if and only if there exist computable reals $a > 0, b$ such that $u^* = au + b$.*

Proof. See Appendix. □

From Theorem 1 one can derive the sufficient direction of a computable vNM representation theorem. That is, we can show that if the agent's preferences are computable continuous, then they can be represented by a computable continuous utility function that they maximize in expectation.

Theorem 2 (Computable Expected Utility)**.** *Under the assumptions of Theorem 1, let* $v : \mathscr{P}(X) \to \mathbb{R}$ *be a computable continuous function satisfying (4) and (5), and let* $u : X \to \mathbb{R}$ *be defined* $u(x) = v(\delta_x)$ *for all* $x \in X$*. Then* u *is a computable continuous function satisfying (1).*

Proof. See Appendix. □

Thus there exists a computable continuous von Neumann-Morgenstern utility function on a computable Polish space X if the preference relation $\preceq$ is itself computable continuous. More generally, the proof shows that given an agent's preference relation we can compute a vNM utility that represents that relation. In this sense we have proved a computable version of (the sufficient direction of) the vNM representation theorem.

Of course, the classical theorem is in fact a biconditional, and the necessary direction is valuable because it shows us that weaker axioms do not suffice. Here we can report only a partial converse to Theorem 2, which relies on the following notion.

Definition 6. A set $K \subseteq X$ is *computably compact* if there is a partial computable function which, given an index for a computable sequence of c.e. open sets $U_0, U_1, \ldots$ in X which covers K, returns a natural number $n \geq 0$ such that $U_0, \ldots, U_n$ covers K.

Theorem 3. *Let* $u : X \to \mathbb{R}$ *be a computable continuous function on a computably compact set* X*. Then the relation* $\preceq$ *defined*

$$P \preceq Q \iff \mathbb{E}_P[u] \leq \mathbb{E}_Q[u]$$

satisfies $(R1^*)$, $(R2)$, *and* $(R3)$.

Proof. See Appendix. □

It is an open question whether this result can be improved by dropping the assumption of computable compactness.

7 Conclusion

We have seen that representations theorems are fundamentally important on any standard interpretation of decision theory. Further, standard decision theory describes a highly idealized agent, one who is not limited in time or space to perform their computations. Work in cognitive science and psychology, however, investigates the kind of reasoning that a highly bounded decision theoretic agent is capable of. I have argued that there is a philosophically rich study lying at the midpoint of these two enterprises: computable decision theory.

I proposed to initiate this study at its foundations via a computable version of the vNM representation theorem. This theorem was motivated by the question: can we compute a vNM utility function that represents an agent's preference relation? Theorem 2 answers this question in the affirmative: the agent's preference relation is sufficient to compute a utility, and if the agent's preference relation is itself computable continuous then the resulting utility is computable continuous. We also proved a partial converse (Theorem 3), but it is an open question whether this can be improved to a full converse of Theorem 2. That said, we have successfully shown that if we have access to an agent's preferences, and if that choice behavior satisfies the vNM axioms, then not only does there exist a vNM utility representing that behavior, but also we can compute that utility function. We are therefore able to use the vNM theorem in a descriptive capacity under very general conditions: for example, we are able to compute a utility function for an agent whose preferences are defined over all possible probability measures on the real numbers.

This project forms a natural starting point for a larger field of study. As was mentioned earlier, there are other representation theorems in decision theory, some of which derive probability in addition to utility. Their scope of application is also wider—the vNM theorem applies only to choice behavior on gambles with known chance distributions. The theorem does not cover choice behavior over options with unknown chances, such as presidential races or stock market predictions. More sophisticated representation theorems such as Savage's can accommodate these cases as well. Thus we might ask whether the Savage representation theorem is similarly computable, in the sense that both the utility *and* probability can be effectively computed from the agent's choice behavior. Future work may explore computable versions of these other representation theorems as well.

A Appendix

Our first step is to prove Theorem 1. To do so we present a sketch of the classical proof of Proposition 2. The proof of our own Theorem 1 follows a parallel structure, so it is informative to work through this proof. We require a few preliminary results. We first show that $\mathscr{D}$ is $\preceq$-order dense in the sense that for any $P \prec Q$ there is $D \in \mathscr{D}$ such that $P \prec D \prec Q$.

Lemma 1. *Let X be a computable Polish space that is moreover a mixture space, and let $\preceq$ be a computable continuous total preorder on X. Then $\mathscr{D}$ is $\preceq$-order dense.*

Proof. If for all $P, Q \in X$, $P \sim Q$, then the lemma is trivial. So let $P, Q \in X$ such that $P \prec Q$. Since $\preceq$ is continuous on X, we have $(P, \rightarrow)$ and $(\leftarrow, Q)$ are open in X. Moreover $(P, \rightarrow) \cap (\leftarrow, Q)$

is nonempty, since for all $\alpha \in [0,1]$, $(\alpha P + (1-\alpha)Q) \in X$, and $P \prec \alpha P + (1-\alpha)Q \prec Q$. Since $\mathscr{D}$ is dense, there is $D \in \mathscr{D}$ such that $D \in (P, \to) \cap (\leftarrow, Q)$. □

Proposition 1. *Suppose $\mathscr{P}$ is a mixture space and $\preceq$ is a binary relation on $\mathscr{P}$. Suppose*

(R1) $\preceq$ is a total preorder;

(R2) if $P \preceq Q$ and $\alpha \in [0,1]$ then $\alpha P + (1-\alpha)R \preceq \alpha Q + (1-\alpha)R$ for all $R \in \mathscr{P}$; and

(R3) if $P \preceq Q \preceq R$ then there exist $\alpha, \beta \in (0,1)$ such that $\alpha P + (1-\alpha)R \preceq Q \preceq \beta P + (1-\beta)R$.

Then if $P \preceq Q$, $Q \preceq R$, and $P \prec R$ then $Q \sim \alpha P + (1-\alpha)R$ for exactly one $\alpha \in [0,1]$;

Proof. See ([13], Theorem 8.3). □

Proposition 2 (Mixture Space Theorem)**.** *Suppose $\mathscr{P}$ is a mixture space and $\preceq$ is a binary relation on $\mathscr{P}$. Then*

(R1) $\preceq$ is a total preorder;

(R2) if $P \preceq Q$ and $\alpha \in [0,1]$ then $\alpha P + (1-\alpha)R \preceq \alpha Q + (1-\alpha)R$ for all $R \in \mathscr{P}$; and

(R3) if $P \preceq Q \preceq R$ then there exist $\alpha, \beta \in (0,1)$ such that $\alpha P + (1-\alpha)R \preceq Q \preceq \beta P + (1-\beta)R$,

if and only if there exists a function $v : \mathscr{P} \to \mathbb{R}$ such that

$$P \preceq Q \iff v(P) \leq v(Q), \text{ and} \tag{4}$$

$$v(\alpha P + (1-\alpha)Q) = \alpha v(P) + (1-\alpha)v(Q). \tag{5}$$

Moreover, if v represents $\preceq$ in the sense of (4) and (5), v' is another representation if and only if $v' = \alpha v + \beta$ for constants $\alpha > 0$ and β.

Proof. We sketch the proof due to ([13], Theorem 8.4). Fix $R, S \in \mathscr{P}(X)$. Let $f(R) = 0$, $f(S) = 1$. Then by Proposition 1, for all $P \in (R, S)$, there is a unique $f(P) \in [0,1]$ such that

$$P \sim [1 - f(P)]R + f(P)S.$$

[13] Theorem 8.4, Part I then shows that for all $P, Q \in (R, S)$ and $\alpha \in [0,1]$,

$$P \prec Q \iff f(P) < f(Q)$$

and

$$f(\alpha P + (a-\alpha)Q) = \alpha f(P) + (1-\alpha)f(Q).$$

(Fishburn, Theorem 8.4 Part II) shows that f can be extended to all of $\mathscr{P}(X)$ as follows. Let $R_1, R_2, S_1, S_2 \in \mathscr{P}(X)$ be such that $(R, S) \subseteq (R_i, S_i)$ for $i = 1, 2$. By the above argument there exist functions f_i^* on (R_i, S_i) satisfying (4) and (5). Moreover they can be linearly transformed into functions f_i such that $f_i(R) = 0$ and $f_i(S) = 1$ for $i = 1, 2$, and this transformation preserves (4) and (5). We then argue that in fact f_1 and f_2 must agree on the intersection of their domains. Given $P \in (R_1, S_1) \cap (R_2, S_2)$ there are three possibilities:

$$\begin{aligned} P \prec R \prec S, &\qquad R \sim (1-\alpha)P + \alpha S \\ R \prec P \prec S, &\qquad P \sim (1-\beta)R + \beta S \\ R \prec S \prec P, &\qquad S \sim (1-\gamma)R + \gamma P \end{aligned}$$

for some $\alpha, \beta, \gamma \in [0, 1]$. It then follows that

$$\begin{aligned} 0 &= (1-\alpha)f_i(P) + \alpha \\ f_i(P) &= \beta \\ 1 &= \gamma f_i(P) \end{aligned}$$

for $i = 1, 2$, so $f_1(P) = f_2(P)$. We let $v(P) = f_i(P)$ for any such function defined on an interval (R_i, S_i) containing (R, S). Then since for any $P \in \mathscr{P}(X)$ there exist R, S such that $P \in (R, S)$, v is defined on all of $\mathscr{P}(X)$ and satisfies (4) and (5). □

One might wonder whether the above proof is already computable—after all, the proof gives a construction of v by fixing a zero and unit and then calibrating $v(P)$ for all other P in terms of the chosen scale. However, the proof relies essentially on Proposition 1, which states only that *there exists* a unique real α which determines the "calibration" of P in terms of the zero and unit. It does not show that this α can be explicitly computed, so we do not know that $v(P)$ can be computed. Indeed the proof of Proposition 1 relies on a continuity argument to show that such a real must exist but does not explicit construct α.

Further, Proposition 2 isn't yet an expected utility representation because v is defined on $\mathscr{P}(X)$, not X, and we have not shown that there is a u such that $P \preceq Q \iff \mathbb{E}_P[u] \leq \mathbb{E}_Q[u]$. The remaining steps are actually quite simple; see e.g. [13], Theorem 10.1.

Fortunately we are able to skip a few steps. We will assume that $\preceq$ is computable continuous; by definition $\preceq$ is continuous. Then the continuity of $\preceq$, $(R2)$, and $(R3)$, by the results of [18] and [14], imply not only that such a utility function $u : X \to \mathbb{R}$ exists, but that it is continuous. Thus our assumptions imply that such a utility function exists, and we need only show that u is computable continuous. To do so we require one final preliminary lemma. It states that, given a function u satisfying (4) and (5), we can compute the members of the countable dense set whose value under u is above or below any given rational.

Lemma 2. *Let X be a computable Polish space, and let $\preceq$ be a computable continuous total preorder on $\mathscr{P}(X)$. Let u be a function satisfying (4) and (5) as given by Proposition 2, with $u(R) = 0$ and $u(S) = 1$ for some $R, S \in \mathscr{D}_{\mathscr{P}(X)}$. Let $q \in \mathbb{Q}$. Then the sets*

$$\Delta_{<q} = \{D \in \mathscr{D} \mid u(D) < q\}$$

and

$$\Delta_{>q} = \{D \in \mathscr{D} \mid u(D) > q\}$$

are c.e.

Proof. There are three cases: $q < 0$, $0 \leq q \leq 1$, or $1 < q$. Suppose $0 \leq q \leq 1$. First note that $u[(1-q)R + qS] = q$, and since $R, S \in \mathscr{D}$, $(1-q)R + qS \in \mathscr{D}$. Search for all $D_i \in \mathscr{D}$ such that $D_i \prec (1-q)R + qS$ using the preferred enumeration. This process is c.e. since $\prec_{\mathscr{D}}$ is c.e. This shows that the set

$$\Delta_{<q} = \{D_i \in \mathscr{D} \mid u(D_i) < q\} = \{D_i \in \mathscr{D} \mid D_i \prec (1-q)R + qS\}$$

is c.e. Similarly the set

$$\Delta_{>q} = \{D_i \in \mathscr{D} \mid u(D_i) > q\} = \{D_i \in \mathscr{D} \mid (1-q)R + qS \prec D_i\}$$

is c.e.

Suppose $q < 0$. Search for $\alpha \in \mathbb{Q} \cap [0,1]$ such that $\frac{-\alpha}{(1-\alpha)} = q$, which is computable since equality is computable on rationals. In particular if $q = \frac{-a}{b}$ with $a, b \in \omega$ then $\alpha = \frac{a}{a-b}$. If for any $P \in \mathscr{P}(X)$, $(1-\alpha)P + \alpha S \sim R$, then $u(P) = q$. Then we may search for all $D_i \in \mathscr{D}$ such that

$$(1-\alpha)D_i + \alpha S \prec R.$$

This process is c.e. since $\prec_{\mathscr{D}}$ is c.e. We thus define the c.e. sets

$$\Delta_{<q} = \{D_i \in \mathscr{D} \mid (1-\alpha)D_i + \alpha S \prec R]\}.$$

and

$$\Delta_{>q} = \{D_i \in \mathscr{D} \mid R \prec (1-\alpha)D_i + \alpha S\}$$

the latter of which is c.e. by a parallel argument.

Finally assume $q > 1$. Let $\alpha = 1/q$. Then define the c.e. sets

$$\Delta_{>q} = \{D_i \in \mathscr{D} \mid S \prec (1-\alpha)R + \alpha D_i\}$$

and

$$\Delta_{<q} = \{D_i \in \mathscr{D} \mid (1-\alpha)R + \alpha D_i \prec S\}.$$

□

Theorem 1 (Computable Mixture Space Theorem)**.** *Let X be a computable Polish space, and let $\mathscr{P}(X)$ be the space of Borel probability measures on X. Assume*

(R1) $\preceq$ is a computable continuous total preorder on $\mathscr{P}(X)$,*

and assume (R2), (R3) hold. Then there exists a computable continuous function $v : \mathscr{P} \to \mathbb{R}$ that satisfies (4) *and* (5).

Proof. Let $R, S \in \mathscr{D}$ such that R, S are not endpoints, and let v be such that $v(R) = 0$, $v(S) = 1$, and v satisfies (4) and (5), as assured by Proposition 2. To show that v is computable continuous it suffices to show that for all rational q, $v^{-1}(-\infty, q), v^{-1}(q, \infty)$ are c.e. open uniformly in q.

We saw above that v must be bounded. Thus note that for any $q \leq \inf v[\mathscr{P}(X)]$, $v^{-1}(-\infty, q) = \emptyset$ and $v^{-1}(q, \infty) = \mathscr{P}(X)$, both of which are c.e. open. Similarly for any $q \geq \sup v[\mathscr{P}(X)]$, $v^{-1}(-\infty, q) = \mathscr{P}(X)$ and $v^{-1}(q, \infty) = \emptyset$.

Thus let $q \in \text{ran}(v)$. By Lemma 2 the sets

$$\Delta_{<q} = \{D_i \in \mathscr{D} \mid v(D) < q\}$$

and

$$\Delta_{>q} = \{D_i \in \mathscr{D} \mid v(D) > q\}$$

are both c.e. uniformly in q.

Given q, let $q_n \to q$ be a uniformly computable fast Cauchy sequence of rationals such that for all n, $q_n < q$. Note that the set

$$\Delta_{q_n < d < q_{n+1}} = \{D_i \in \mathscr{D} \mid q_n < v(D) < q_{n+1}\}$$

is an intersection of c.e. sets and hence is c.e. Define a total computable function $m : \omega \to \omega$ by letting $m(n)$ be the index i of the first $D_i \in \mathscr{D}$ enumerated into the set $\Delta_{q_n < d < q_{n+1}}$. For any q and any $D_{m(n)}$ we have $(\leftarrow, D_{m(n)}) \subseteq v^{-1}(-\infty, q)$; thus

$$\bigcup_{n\in\omega}(\leftarrow, D_{m(n)}) \subseteq v^{-1}(-\infty, q).$$

Furthermore $\bigcup_{n\in\omega}(\leftarrow, D_{m(n)})$ is a computable union of c.e. open sets and hence is c.e. open.

We now show that $v^{-1}(-\infty, q) \subseteq \bigcup_{n\in\omega}(\leftarrow, D_{m(n)})$. Let $Q \in \mathscr{P}(X)$ with $v(Q) < q$. There is n such that $2^{-n} \leq |v(Q) - q|$, in which case $Q \in (\leftarrow, D_{m(n)})$ by definition of m. Therefore $v^{-1}(-\infty, q) = \bigcup_{n\in\omega}(\leftarrow, D_{m(n)})$, a c.e. open set uniformly in q. A parallel argument establishes that there is a total computable function m' such that $v^{-1}(q, \infty) = \bigcup_{n\in\omega}(D_{m'(n)}, \rightarrow)$. □

Corollary 2. *The function $u : X \to \mathbb{R}$ constructed in Theorem 1 is unique up to computable positive linear transformation; that is, $u^* : X \to \mathbb{R}$ satisfies (4) and (5) if and only if there exist computable reals $a > 0, b$ such that $u^* = au + b$.*

Proof. The necessary direction is obvious. For the sufficient direction there are two cases. If u is constant then so is u^*, and so $u^* = u + (c - c')$ where $u(P) = c, u^*(P) = c'$ for all $P \in \mathscr{P}$. Since u and u^* are computable continuous, c and c' are computable reals. Otherwise let $R, S \in \mathscr{D}$ with $R \prec S$. Then since u and u^* are computable continuous, we have that $u(R), u(S), u^*(R), u^*(S)$ are computable reals. By the classical proof of uniqueness (see [13], Theorem 8.4) we have for all $P \in \mathscr{P}$,

$$u^*(P) = \frac{u^*(S) - u^*(R)}{u(S) - u(R)} u(P) + u^*(R) - u(R)\frac{u^*(S) - u^*(R)}{u(S) - u(R)}.$$

Then let

$$a = \frac{u^*(S) - u^*(R)}{u(S) - u(R)}$$
$$b = u^*(R) - u(R)\frac{u^*(S) - u^*(R)}{u(S) - u(R)}$$

which are computable reals. □

If in addition to satisfying axioms $(R1), (R2)$, and $(R3)$, the relation $\preceq$ is required to be continuous, then there is a utility function u' that both satisfies (1) *and* is continuous (see [39] Theorem IV.2.7). Therefore the assumptions of Theorem 1 also imply the existence of such a u. The standard trick to move from the function $u : \mathscr{P}(X) \to \mathbb{R}$ as given in Theorem 1 to the desired function $u' : X \to \mathbb{R}$ satisfying (1) is to set $u'(x) = u(\delta_x)$, where δ_x is the Dirac measure concentrated on the point x; see ([13], Theorem 10.1). Our main theorem shows that u', so defined, is a computable continuous function satisfying (1).

Theorem 2 (Computable Expected Utility)**.** *Under the assumptions of Theorem 1, let* $v : \mathscr{P}(X) \to \mathbb{R}$ *be a computable continuous function satisfying (4) and (5), and let* $u : X \to \mathbb{R}$ *be defined* $u(x) = v(\delta_x)$ *for all* $x \in X$*. Then* u *is a computable continuous function satisfying (1).*

Proof. By [13], Theorem 10.1, u satisfies (1), so it suffices to show that u is computable continuous. Thus we want to show that the sets $u^{-1}(-\infty, q)$ and $u^{-1}(q, \infty)$ are c.e. open uniformly in $q \in \mathbb{Q}$. By definition of u we have

$$x \in u^{-1}(-\infty, q) \iff v(\delta_x) < q.$$

Now, for any Dirac measures δ_x, δ_y for $x, y \in X$, the Prokhorov metric $d_P(\delta_x, \delta_y)$ satisfies

$$d_P(\delta_x, \delta_y) < \epsilon \iff \delta_x(U) - \epsilon < \delta_y(U^\epsilon)$$

for all $U \in \text{supp}(\delta_x)$. But since X is a separable metric space we have that $\text{supp}(\delta_x) = \{x\}$. Thus we have

$$d_P(\delta_x, \delta_y) < \epsilon \iff \delta_x(\{x\}) - \epsilon < \delta_y(B_X(x, \epsilon)),$$

where $B_X(x, \epsilon) \subseteq X$. The right-hand inequality holds iff either $y \in B_X(x, \epsilon)$ or $d(x, y) \geq 1$, in which case $d_P(\delta_x, \delta_y) = 1$. In other words,

$$\delta_y \in B_{\mathscr{P}(X)}(\delta_x, \epsilon) \iff y \in B_X(x, \epsilon) \tag{6}$$

for all $x, y \in X$ with $d(x, y) < 1$.

Consider the set $v^{-1}(-\infty, q)$. Since v is computable continuous, there is a partial computable function that takes $q \in \mathbb{Q}$ and returns an index for a c.e. set I such that

$$v^{-1}(-\infty, q) = \bigcup_{(i,r) \in I} B_{\mathscr{P}(X)}(\delta_{s_i}, r) \tag{7}$$

where $s_i \in \mathscr{D}$. In particular $v^{-1}(-\infty, q)$ can be written so that r is uniformly less than 1 for all $(i, r) \in I$. Therefore by (6) and (7),

$$u^{-1}(-\infty, q) = \bigcup_{(i,r) \in I} B_X(s_i, r),$$

a c.e. open set uniformly in q. A similar argument shows that $u^{-1}(q, \infty)$ is c.e. open uniformly in q. □

We now turn to proving the partial converse to Theorem 2. We introduced computable compactness because it is known that the suprema and infima of computable continuous functions are themselves computable on computably compact domains.

Proposition 3. *If X is computably compact and $f : X \to \mathbb{R}$ is computable continuous, then both $\sup_{x \in X} f(x)$ and $\inf_{x \in X} f(x)$ are computable reals.*

As a final preliminary step we present a result due to Hoyrup and Rojas ([21]) on the computability of the integral.

Proposition 4. *Let $f : X \to \mathbb{R}$ be a computable continuous function with a computable bound M. Then the integral $\mu \mapsto \int f \, d\mu$ is a computable continuous map.*

Proof. See [21], Corollary 4.3.2. □

Theorem 3. *Let $u : X \to \mathbb{R}$ be a computable continuous function on a computably compact set X. Then the relation $\preceq$ defined*

$$P \preceq Q \iff \mathbb{E}_P[u] \leq \mathbb{E}_Q[u]$$

satisfies $(R1^*)$, $(R2)$, and $(R3)$.

Proof. $(R2)$ and $(R3)$ follow from the classic proof. To see $(R1^*)$, note that by Proposition 3, u has a computable bound $M = \max\{\sup_{x \in X} u, \inf_{x \in X} u\}$. Therefore by Proposition 4, the integral $v : P \mapsto \int u \, dP$ is a computable continuous function. Fix $D \in \mathscr{D}$, and note that $\mathbb{E}_D[u] = \alpha_D$ is a computable real. Then the sets $(\leftarrow, D) = v^{-1}(-\infty, \alpha_D)$ and $(D, \to) = v^{-1}(\alpha_D, \infty)$ are c.e. open, establishing $(R1^*)$. □

References

[1] F. J. Anscombe and R. Aumann. A definition of subjective probability. *The Annals of Mathematical Statistics*, 34(1):199–205, 1963.

[2] K. J. Arrow. The theory of risk aversion. In *Essays in the Theory of Risk Bearing*. North Holland, 1965.

[3] José Luis Bermúdez. *Decision Theory and Rationality*. Oxford University Press, 2009.

[4] Patrick Billingsley. *Convergence of Probability Measures*. Wiley, 2nd edition, 1999.

[5] Vladimir I. Bogachev. *Weak Convergence of Measures*. American Mathematical Society, 2018.

[6] Vasco Brattka, Peter Hertling, and Klaus Weihrauch. A tutorial on computable analysis. In *New Computational Paradigms*. Springer, 2008.

[7] R. A. Briggs. Normative Theories of Rational Choice: Expected Utility. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2023 edition, 2023.

[8] Lara Buchak. Normative Theories of Rational Choice: Rivals to Expected Utility. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2022 edition, 2022.

[9] Laura Buchak. *Risk and Rationality*. Oxford University Press, 2013.

[10] Laura Buchak. Decision theory. In Alan Hájek and Christopher Hitchcock, editors, *The Oxford Handbook of Probability and Philosophy*, pages 789–811. Oxford University Press, 2016.

[11] Jennifer R. Carr. Why ideal epistemology? *Mind*, 131(524):1131–1162, 2022.

[12] David Danks and Frederick Eberhardt. Keeping bayesian models rational: The need for an account of algorithmic rationality. *Behavioral and Brain Sciences*, 34(4):197–197, 2011.

[13] Peter Fishburn. *Utility Theory for Decision Making*. Robert E. Krieger Publishing Company, 1970.

[14] Lucien Foldes. Expected utility and continuity. *The Review of Economic Studies*, 39(4):407–421, 1972.

[15] M. Friedman and L. J. Savage. The utility analysis of choices involving risk. *Journal of Political Economy*, 56:279–304, 1948.

[16] Milton Friedman and L. J. Savage. The expected-utility hypothesis and the measurability of utility. *Journal of Political Economy*, 60(6):463–474, 1952.

[17] Samuel J. Gershman, Eric J. Horvitz, and Joshua B. Tenenbaum. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349(6245):273–278, 2015.

[18] Jean-Michel Grandmont. Continuity properties of expected utility. *Journal of Economic Theory*, 4:45–57, 1972.

[19] Gilbert Harman. *Change in View*. MIT Press, 1986.

[20] I. N. Herstein and John Milnor. An axiomatic approach to measurable utility. *Econometrica*, 21(2):291–297, 1953.

[21] Mathieu Hoyrup and Cristóbal Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. *Information and Computation*, 207(7):830–847, 2009.

[22] Thomas Icard. Why be random? *Mind*, 130(517):111–139, 2021.

[23] Thomas Icard. Bayes, bounds, and rational analysis. *Philosophy of Science*, 85(1):79–101, 2022.

[24] James M. Joyce. *The Foundations of Causal Decision Theory*. Cambridge University Press, 1999.

[25] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

[26] Saira Khan. Rational preference in transformative experiences. *Synthese*, 199:6715–6732, 2021.

[27] David Kreps. *Notes on the Theory of Choice*. Westview Press, 1988.

[28] David Lewis. Radical interpretation. *Synthese*, 27(3/4):331–344, 1974.

[29] David Lewis. Causal decision theory. *Australasian Journal of Philosophy*, 59(1):5–30, 1981.

[30] Richard L. Lewis, Andrew Howes, and Satinder Singh. Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science*, 6:279–311, 2014.

[31] Mike Oaksford and Nick Chater. *Rational Models of Cognition*. Oxford University Press, 1999.

[32] Samir Okasha. On the interpretation of decision theory. *Economics and Philosophy*, 32:409–433, 2016.

[33] John W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1–2):122–136, 1964.

[34] Frank P. Ramsey. Truth and probability. In R. Braithwaite, editor, *The Foundations of Mathematics and Other Logical Essays*, pages 156–198. Kegan Paul, 1931.

[35] Leonard J. Savage. *The Foundations of Statistics*. Wiley & Sons, 1954.

[36] Teddy Seidenfeld, Mark J. Schervish, and Joseph B. Kadane. A representation of partially ordered preferences. *The Annals of Statistics*, 23(6):2168–2217, 1995.

[37] Herbert Simon. *Reason in Human Affairs*. Stanford University Press, 1983.

[38] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[39] Peter P. Wakker. *Additive Representations of Preferences*. Springer, 1989.

[40] Klaus Weihrauch. *Computable Analysis: An Introduction*. Springer, 2000.

[41] Paul Weirich. *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*. Oxford University Press, 2004.

[42] L. Zynda. Representation theorems and realism about degrees of belief. *Philosophy of Science*, 67(1):45–69, 2000.