Constructive Decision Theory

Lawrence Blume[a,c,d], David Easley[a] and Joseph Y. Halpern[b]

[a] Department of Economics, Cornell University
[b] Department of Computer Science, Cornell University
[c] Institute for Advanced Studies, Vienna
[d] The Santa Fe Institute

June 16, 2009

Supported in part by NSF under grants CTC-0208535, ITR-0325453, and IIS-0534064, by ONR under grants N00014-00-1-03-41 and N00014-01-10-511, and by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the ONR under grant N00014-01-1-0795. An extended abstract of this paper (Blume, Easley, and Halpern 2006) appeared in the *Tenth International Conference on Principles of Knowledge Representation and Reasoning*. This paper was written for and first presented at a conference on decision theory organized by Karl Vind in May 2004, shortly before his death. As a scholar and a colleague, Karl has left us all much to be grateful for. We are particularly grateful to him for pushing us to deliver a paper for this conference, and then to generously comment on it despite his ill health. We are also grateful to several individuals for comments and to the many seminar audiences who have listened to and commented on this paper.

Abstract: In most contemporary approaches to decision making, a decision problem is described by a sets of states and set of outcomes, and a rich set of *acts*, which are functions from states to outcomes over which the decision maker (DM) has preferences. Most interesting decision problems, however, do not come with a state space and an outcome space. Indeed, in complex problems it is often far from clear what the state and outcome spaces would be. We present an alternative foundation for decision making, in which the primitive objects of choice are *syntactic programs*. A representation theorem is proved in the spirit of standard representation theorems, showing that if the DM's preference relation on objects of choice satisfies appropriate axioms, then there exist a set S of states, a set O of outcomes, a way of interpreting the objects of choice as functions from S to O, a probability on S, and a utility function on O, such that the DM prefers choice a to choice b if and only if the expected utility of a is higher than that of b. Thus, the state space and outcome space are subjective, just like the probability and utility; they are not part of the description of the problem. In principle, a modeler can test for SEU behavior without having access to states or outcomes. We illustrate the power of our approach by showing that it can capture decision makers who are subject to framing effects.

Correspondent:
Professor Lawrence Blume
Department of Economics
Uris Hall
Cornell University
Ithaca NY 14850

In memoriam Karl Vind.

1 Introduction

Most models of decisionmaking under uncertainty describe a decision environment with a set of states and a set of outcomes. Objects of choice are *acts*, functions from states to outcomes. The decision maker (DM) holds a preference relation on the set of all such functions. Representation theorems characterize those preference relations with utility functions on acts that separate (more or less) tastes on outcomes from beliefs on states. The canonical example is Savage's (1954) characterization of those preference relations that have a subjective expected utility (SEU) representation: Acts are ranked by the expectation of a utility payoff on their outcomes with respect to a probability distribution on states. *Choquet expected utility* (Schmeidler 1989) maintains the separation between tastes and beliefs, but does not require that beliefs be represented by an additive measure. Tversky and Kahneman's (1992) *cumulative prospect theory* relaxes the taste-belief separation by assessing gains and losses with different belief measures; Wakker and Tversky (1993) discuss generalizations of SEU from this point of view. Modern attempts to represent *ambiguity* in choice theory relax both the meaning of likelihood and the separation of tastes and beliefs that characterize SEU. All of these generalizations of SEU, however, maintain the state-outcome-act description of objects of choice and, moreover, take this description of choice problems as being given prior to the consideration of any preference notion.

We, on the other hand, follow Ellsberg (2001) in locating the source of ambiguity in the description of the problem. For Savage (1954, p. 9), the *world* is 'the object about which the person is concerned' and a *state* of the world is 'a description of the world, leaving no relevant aspect undescribed.' But what are the 'relevant' descriptors of the world? Choices do not come equipped with states. Instead they are typically objects described by their manner of realization, such as 'buy 100 shares of IBM' or 'leave the money in the bank,' 'attack Iraq,' or 'continue to negotiate.' In Savage's account (1954, sec. 2.3) it is clear that the DM 'constructs the states' in contemplating the decision problem. In fact, his discussion of the rotten egg foreshadows this process. Subsequently, traditional decision theory has come to assume that states are given as part of the description of the decision problem. We suppose instead that states are constructed by the DM in the course of deliberating about questions such as 'How is choice A different from choice B?' and 'In what circumstances will choice A turn out better than choice B?'. These same considerations apply (although here Savage may disagree) to outcomes. This point has been forcefully made by Weibull (2004).

There are numerous papers in the literature that raise issues with the state-space approach of Savage or that derive a subjective state space. Machina (2006) surveys the standard approach and illustrates many difficulties with the theory and with its uses. These difficulties include the ubiquitous ambiguity over whether the theory is meant to be descriptive or normative, whether states are exogenous or constructed by the DM, whether states are external to the DM, and whether they are measurable or not. Kreps (1992) and Dekel, Lipman, and Rustichini (2001) use a menu choice model to deal with unforeseen contingencies—an inability of the DM to list all possible states of the world. They derive a subjective state space that represents possible preference orders over elements of the menu chosen by the DM. Ghirardato (2001) takes an alternative approach to unforeseen contingencies and models acts as correspondences from a state space to outcomes. Gilboa and Schmeidler (2004) and Karni (2006) raise objections to the state space that are similar to ours and develop decision theories without a state space. Both papers derive subjective probabilities directly on outcomes. Ahn (2008) also develops a theory without a state space; in his theory, the DM chooses over sets of lotteries over consequences. Ahn and Ergin (2007) allow for the possibility that there may be different descriptions of a particular event, and use this possibility to capture framing. For them, a 'description' is a partition of the state space. They provide an axiomatic foundation for decision making in this framework, built on Tversky and Koehler's (1994) notion of *support theory*.

Our approach differs significantly from these mentioned above. The inspiration for our approach is the observation that objects of choice in an uncertain world have some structure to them. Individuals choose among some simple actions: 'buy 100 shares of IBM' or 'attack Iraq'. But they also perform various tests on the world and make choices contingent upon the outcome of these tests: 'If the stock broker recommends buying IBM, then buy 100 shares of IBM; otherwise buy 100 shares of Google.' These tests are written in a fixed language (which we assume is part of the description of the decision problem, just as Savage assumed that states were part of the description of the decision problem). The language is how the DM describes the world. We formalize this viewpoint by taking the objects of choice to be (syntactic) programs in a programming language.

The programming language is very simple—we use it just to illustrate our ideas. Critically, it includes tests (in the context of **if** . . . **then** . . . **else** statements). These tests involve syntactic descriptions of the events in the world, and allow us to distinguish events from (syntactic) descriptions of events. In particular, there can be two different descriptions that, intuitively, describe the same event from the point of view of the modeler but may describe different events from the point of view of the decision maker. Among other things, this enables us to capture framing effects in our framework, without requiring states as Ahn and Ergin (2007) do, and provides a way of dealing with resource-bounded reasoners.

In general, we do not include outcomes as part of the description of the decision problem; both states and outcomes are part of the DM's (subjective) representation of the problem. We assume that the DM has a weak preference relation on the objects of choice; we do not require the preference relation to be complete. The set of acts for a decision problem is potentially huge, and may contain acts that will never be considered by the DM. While we believe that empirical validity requires considering partial orders, there are also theoretical reasons for considering partial orders. Our representation theorems for partial orders require a set of probabilities and utility functions (where often one of the sets can be taken to be a singleton). Schmeidler (1989, p. 572) observes that using a set of probability distributions can be taken as a measure of a DM's lack of confidence in her likelihood assessments. Similarly, a set of utilities can be interpreted as a lack of confidence in her taste assessments (perhaps because she has not had time to think them through carefully).

The rest of this paper is organized as follows. We begin the next section with a description of the syntactic programs that we take as our objects of choice, discuss several interpretations of the model, and show how syntactic programs can be interpreted as Savage acts. In Section 3, we present our assumptions on preferences The key postulate is an analogue of Krantz et al.'s (1971) *cancellation axiom*. In Section 4 we present our representation theorems for decision problems with subjective outcomes and those with objective outcomes. Section 5 discusses how our framework can model boundedly rational reasoning. In Section 6 we discuss how updating works for new information about the external world as well as for new information about preferences. Section 7 concludes.

2 Describing Choices

We begin by describing the language of tests, and then use this language to construct our syntactic objects of choice. We then use the language of tests to describe theories of the world. We show how framing problems can be understood as 'odd' theories held by decision makers.

2.1 Languages for tests and choices

A *primitive test* is a yes/no question about the world, such as, 'IBM's price-earnings ratio is 5', 'the economy will be strong next year' and 'the moon is in the seventh house'. We assume a finite set T_0 of primitive tests. The set T of tests is constructed by closing the set of primitive tests under conjunction and negation. That is, T is the smallest set such that $T_0 \subseteq T$, and if t_1 and t_2 are in T, so is $t_1 \wedge t_2$ and $\neg t_1$. Thus, the language of tests is just a propositional logic whose atomic propositions are the elements of T_0.

We consider two languages for choices. In both cases, we begin with a finite set $\mathcal{A}_0$ of *primitive choices*. These may be objects such as 'buy 100 shares of IBM' or 'buy \$10,000 worth of bonds'. The interpretation of these acts is tightly bound to the decision problem being modeled. The first language simply closes off $\mathcal{A}_0$ under **if** ... **then** ... **else**. By this we mean that if t is a test in T and a and b are choices in $\mathcal{A}$, then **if** t **then** a **else** b is also a choice in $\mathcal{A}$. When we need to be clear about which T_0 and A_0 were used to construct $\mathcal{A}$, we will write $\mathcal{A}_{\mathcal{A}_0,T_0}$. Note that $\mathcal{A}$ allows nesting, so that **if** t_1 **then** a **else** (**if** t_2 **then** b **else** c) is also a choice.

The second languages closes off $\mathcal{A}_0$ with **if** ... **then** ... **else** and randomization. That is, we assume that objective probabilities are available, and require that for any $0 \leq r \leq 1$, if a and b are choices, so is $ra + (1-r)b$. Randomization and **if** ... **then** ... **else** can be nested in arbitrary fashion. We call this language $\mathcal{A}^+$ ($\mathcal{A}^+_{\mathcal{A}_0,T_0}$ when necessary).

Tests in T are elements of discourse about the world. They could be events upon which choice is contingent: If the noon price of Google stock today is below \$600, then buy 100 shares, else buy none. More generally, tests in T are part of the DM's description of the decision problem, just as states are part of the description of the decision problem in Savage's framework. However, elements of T need not be complete descriptions of the relevant world, and therefore may not correspond to Savage's states. When we construct state spaces, elements of T will clearly play a role in defining states, but, for some of our representation theorems, states cannot be constructed out of elements of T alone. Additional information in states is needed for both incompleteness of preferences and when the outcome space is taken to be objective or exogenously given.

The choices in $\mathcal{A}$ and $\mathcal{A}^+$ are *syntactic* objects; strings of symbols. They can be given *semantics*—that is, they can be interpreted—in a number of ways. For most of this paper we focus on one particular way of interpreting them that lets us connect them to Savage acts, but we believe that other semantic approaches will also prove useful (see Section 7). The first step in viewing choices as Savage acts is to construct a state space S, and to interpret the tests as events (subsets of S). With this semantics for tests, we can then construct, for the state space S and a given outcome space O, a function ρ_{SO} that associates with each choice a a Savage act $\rho_{SO}(a)$, that is, a function from S to O. Given a state space S, these constructions work as follows:

Definition 1. *A* test interpretation π_S *for the state space* S *is a function associating with each test a subset of* S*. An interpretation is* standard *if it interprets* $\neg$ *and* $\wedge$ *in the usual way; that is*

- $\pi_S(t_1 \wedge t_2) = \pi_S(t_1) \cap \pi_S(t_2)$
- $\pi_S(\neg t) = S - \pi_S(t)$.

Intuitively, $\pi_S(t)$ is the set of states where t is true. We will allow for *nonstandard interpretations*. These are interpretations in which, for some test t, there may be some state where neither t nor $\neg t$ is true; that is, there may be some state in neither $\pi_S(t)$ nor $\pi_S(\neg t)$); similarly, there may be some state where both t and $\neg t$ are true. Such nonstandard interpretations are essentially what philosophers call 'impossible possible worlds' (Rantala 1982); they have also been used in game theory for modeling resource-bounded reasoning (Lipman 1999). A standard interpretation is completely determined by its behavior on primitive tests. This is not true of nonstandard interpretations. All test interpretations are assumed to be standard until Section 5. There we motivate nonstandard interpretations, and show how our results can be modified to hold even with them.

Definition 2. *A* choice interpretation ρ_{SO} *for the state space* S *and outcome space* O *assigns to each choice* $a \in \mathcal{A}$ *a (Savage) act, that is, a function* $\rho_{SO}(a) : S \to O$.

Choice interpretations are constructed as follows: Let $\rho^0_{SO} : \mathcal{A}_0 \to O^S$ be a choice interpretation for primitive choices, which assigns to each $a_o \in \mathcal{A}_0$ a function from $S \to O$. We extend ρ^0_{SO} to a function mapping all choices in $\mathcal{A}$ to functions from S to O by induction on structure, by defining

$$\rho_{SO}(\textbf{if } t \textbf{ then } a_1 \textbf{ else } a_2)(s) = \begin{cases} \rho_{SO}(a_1)(s) & \text{if } s \in \pi_S(t) \\ \rho_{SO}(a_2)(s) & \text{if } s \notin \pi_S(t). \end{cases} \tag{1}$$

This semantics captures the idea of contingent choices; that, in the choice **if** t **then** a_1 **else** a_2, the realization of a_1 is contingent upon t, while a_2 is contingent upon 'not t'. Of course, a_1 and a_2 could themselves be compound acts.

Extending the semantics to the language $\mathcal{A}^+$, given S, O, and π_S, requires us to associate with each choice a an Anscombe-Aumann (AA) act (Anscombe and Aumann 1963), that is, a function from S to probability measures on O. Let $\Delta(O)$ denote the set of probability measures on O and let $\Delta^*(O)$ be the subset of $\Delta(O)$ consisting of the probability measures that put probability one on an outcome. Let $\rho^0_{SO} : \mathcal{A}_0 \to \Delta^*(O)^S$ be a choice interpretation for primitive choices that assigns to each $a_o \in \mathcal{A}_0$ a function from $S \to \Delta^*(O)$. Now we can extend ρ^0_{SO} by induction on structure to all of $\mathcal{A}^+$ in the obvious way. For **if** ... **then** ... **else** choices we use (1); to deal with randomization, define

$$\rho_{SO}(ra_1 + (1-r)a_2)(s) = r\rho_{SO}(a_1)(s) + (1-r)\rho_{SO}(a_2)(s).$$

That is, the distribution $\rho_{SO}(ra_1 + (1-r)a_2)(s)$ is the obvious mixture of the distributions $\rho_{SO}(a_1)(s)$ and $\rho_{SO}(a_2)(s)$. Note that we require ρ_{SO} to associate with each primitive choice in each state a single outcome (technically, a distribution that assigns probability 1 to a single outcome), rather than an arbitrary distribution over outcomes. So primitive choices are interpreted as Savage acts, and more general choices, which are formed by taking objective mixtures of choices, are interpreted as AA acts. This choice is largely a matter of taste. We would get similar representation theorems even if we allowed ρ^0_{SO} to be an arbitrary function from $\mathcal{A}$ to $\Delta(O)^S$. However, this choice does matter for our interpretation of the results; see Example 11 for further discussion of this issue.

2.2 Framing and equivalence

Framing problems appear when a DM solves inconsistently two decision problems that are designed by the modeler to be equivalent or that are obviously similar after recognizing an equivalence. The fact that choices are syntactic objects allows us to capture framing effects.

Example 1. Consider the following well-known example of the effects of framing, due to McNeil et al. (1982). DMs are asked to choose between surgery or radiation therapy as a treatment for lung cancer. The problem is framed in two ways. In the what is called the *survival frame*, DMs are told that, of 100 people having surgery, 90 live through the post-operative period, 68 are alive at the end of the first year, and 34 are alive at the end of five years; and of 100 people have radiation therapy, all live through the treatment, 77 are alive at the end of the first year, and 22 are alive at the end of five years. In the *mortality frame*, DMs are told that of 100 people having surgery, 10 die during the post-operative period, 32 die by the end of the first year, and 66 die by the end of five years; and of 100 people having radiation therapy, none die during the treatment, 23 die by the end of the first year, and 78 die by the end of five years. Inspection shows that the outcomes are equivalent in the two frames—90 of 100 people living is the same as 10 out of 100 dying, and so on. Although one might have expected the two groups to respond to the data in similar fashion, this was not the case. While only 18% of DMs prefer radiation therapy in the survival frame, the number goes up to 44% in the mortality frame.

We can represent this example in our framework as follows. We assume that we have the following tests:

- RT, which intuitively represents '100 people have radiation therapy';
- S, which intuitively represents '100 people have surgery';
- $L_i(k)$, for $i = 0, 1, 5$ and $k = 0, \dots, 100$, which intuitively represents that k out of 100 people live through the post-operative period (if $i = 0$), are alive after the first year (if $i = 1$), and are alive after five years (if $i = 5$);
- $D_i(k)$, for $i = 0, 1, 5$ and $k = 0, \dots, 100$, which is like $L_i(k)$, except 'live/alive' are replaced by 'die/dead'.

In addition, we assume that we have primitive programs a_S and a_R that represent 'perform surgery' and 'perform radiation theory'. With these tests, we can characterize the description of the survival frame by the following test t_1:

$$(S \Rightarrow L_0(90) \wedge L_1(68) \wedge L_5(34)) \wedge (RT \Rightarrow L_0(100) \wedge L_1(77) \wedge L_5(22)),$$

(where, as usual, $t \Rightarrow t'$ is an abbreviation for $\neg(t \wedge \neg t')$); similarly, the mortality frame is characterized by the following test t_2:

$$(S \Rightarrow D_0(10) \wedge D_1(32) \wedge D_5(66)) \wedge (RT \Rightarrow D_0(0) \wedge D_1(23) \wedge D_5(78)).$$

The choices offered in the McNeil et al. experiment can be viewed as conditional choices: what would a DM do conditional on t_1 (resp., t_2) being true. Using ideas from Savage, we can capture the survival frame as a decision problem with the following two choices:

if t_1 **then** a_S **else** a, and
if t_1 **then** a_R **else** a,

where a is an arbitrary choice. Intuitively, comparing these choices forces the DM to consider his preferences between a_S and a_R conditional on the test, since the outcome in these two choices is the same if the test does not hold. Similarly, the mortality frame amounts to a decision problem with the analogous choices with t_1 replaced by t_2.

There is nothing in our framework that forces a DM to identify the tests t_1 and t_2; the tests $L_i(k)$ and $D_i(100-k)$ a priori are completely independent, even if the problem statement suggests that they should be equivalent. Hence there is no reason for a DM to identify the choices **if** t_1 **then** a_S **else** a and **if** t_2 **then** a_S **else** a. As a consequence, as we shall see, it is perfectly consistent with our axioms that a DM has the preferences **if** t_1 **then** a_S **else** $a \succ$ **if** t_1 **then** a_R **else** a and **if** t_2 **then** a_R **else** $a \succ$ **if** t_2 **then** a_S **else** a. ∎

We view it as a feature of our framework that it can capture this framing example for what we view as the right reason: the fact that DMs do not necessarily identify $L_i(k)$ and $D_i(100-k)$. Nevertheless, we would also like to be able to capture the fact that more sophisticated DMs do recognize the equivalence of these tests. We can do this by associating with a DM her understanding of the relationship between tests. For example, a sophisticated DM might understand that $L_i(k) \Leftrightarrow D_i(100-k)$, for $i = 0, 1, 5$ and $k = 1, \ldots, 100$. Formally, we add to the description of a decision problem a *theory*, that is, a set AX $\subseteq T$ of tests. Elements of the theory are called *axioms*.

Definition 3. *A test interpretation π_S for the state space S* respects *a theory AX iff for all $t \in$ AX, $\pi_S(t) = S$.*

A theory represents the DM's view of the world; those tests he takes to be axiomatic (in its plain sense). Different people may, however, disagree about what they take to be obviously true of the world. Many people will assume that the sun will rise tomorrow. Others, like Laplace, will consider the possibility that it will not.

Choices a and b are *equivalent with respect to a set Π of test interpretations* if, no matter what interpretation $\pi \in \Pi$ is used, they are interpreted as the same function. For example, in any standard interpretation, **if** t **then** a **else** b is equivalent to **if** $\neg\neg t$ **then** a **else** b; no matter what the test t and choices a and b are, these two choices have the same input-output semantics. The results of the McNeil et al. experiment discussed in Example 1 can be interpreted in our language as a failure by some DMs to have a theory that makes tests stated in terms of mortality data or survival data semantically equivalent. This then allows choices, such as the choice of surgery or radiation therapy given these tests, to be not seen as equivalent by the DM. Thus, it is not surprising that such DMs are not indifferent between these choices.

Definition 4. *For a set Π of test interpretations, choices a and b are Π-equivalent, denoted $a \equiv_\Pi b$, if for all state spaces S, outcomes O, test interpretations $\pi_S \in \Pi$, and choice interpretations ρ_{SO}, $\rho_{SO}(a) = \rho_{SO}(b)$.*

Denote by Π_{AX} the set of all standard interpretations that respect theory AX. Then Π_{AX}-equivalent a and b are said to be AX-equivalent, and we write $a \equiv_{AX} b$. Note that equivalence is defined relative to a given set Π of interpretations. Two choices may be equivalent with respect to the set of all standard interpretations that hold a particular test t to be true, but not equivalent to the larger set of all standard test interpretations.

3 The Axioms

This section lays out our basic assumptions on preferences. Since our basic framework allows for preferences only on discrete sets of objects, we cannot use conventional independence axioms. Instead, we use cancellation. Cancellation axioms are not well known, so we use this opportunity to derive some connections between cancellation and more familiar preference properties.

3.1 Preferences

We assume that the DM has a weak preference relation $\succeq$ on a subset C of the sets $\mathcal{A}$ (resp., $\mathcal{A}^+$) of non-randomized (resp., randomized) acts. This weak preference relation has the usual interpretation of 'at least as good as'. We take $a \succ b$ to be an abbreviation for $a \succeq b$ and $b \not\succeq a$, even if $\succeq$ is not complete. We prove various representation theorems that depend upon the language, and upon whether outcomes are taken to be given or not. The engines of our analysis are various cancellation axioms, which are the subject of the next section. At some points in our analysis we consider complete preferences:

A1. The preference relation $\succeq$ is complete.

The completeness axiom has often been defended by the claim that 'people, in the end, make choices.' Nonetheless, from the outset of modern decision theory, completeness has been regarded as a problem. Savage (1954, Section 2.6) discusses the difficulties involved in distinguishing between indifference and incompleteness. He concludes by choosing to work with the relationship he describes with the symbol $\leqq \cdot$, later abbreviated as $\leq$, which he interprets as 'is not preferred to'. The justification of completeness for the 'is not preferred to' relationship is anti-symmetry of strict preference. Savage, Aumann (1962), Bewley (2002) and Mandler (2001) argue against completeness as a requirement of rationality. Eliaz and Ok (2006) have argued that rational choice theory with incomplete preferences is consistent with preference reversals. In our view, incompleteness is an important expression of ambiguity in its plain meaning (rather than as a synonym for a non-additive representation of likelihood). There are many reasons why a comparison between two objects of choice may fail to be resolved: obscurity or indistinctness of their properties, lack of time for or excessive cost of computation, the incomplete enumeration of a choice set, and so forth. We recognize indecisiveness in ourselves and others, so it would seem strange not to allow for it in any theory of preferences that purports to describe tastes (as opposed to a theory which purports to characterize consistent choice).

3.2 Cancellation

Axioms such as the independence axiom or the sure thing principle are required to get the requisite linearity for an SEU representation. But in their usual form these axioms cannot be stated in our framework as they place restrictions on preference relations over acts and we do not have acts. Moreover, some of our representation theorems apply to finite sets of acts, while the usual statement of mixture axioms requires a mixture space of acts. For us, the role of these axioms is performed by the *cancellation axiom*, which we now describe. Although simple versions of the cancellation axiom have appeared in the literature (e.g. Scott (1964) and Krantz, Luce, Suppes, and Tversky (1971)), it is nonetheless not well known, and so before turning to our framework we briefly explore some of its implications in more familiar settings. Nonetheless, some of the results here are new; in particular, the results on cancellation for partial orders. These will be needed for proofs in the appendix.

Let C denote a set of choices and $\succeq$ a preference relation on C. We use the following notation: Suppose $\langle a_1, \ldots, a_n \rangle$ and $\langle b_1, \ldots, b_n \rangle$ are sequences of elements of C. If for all $c \in C$, $\#\{j : a_j = c\} = \#\{j : b_j = c\}$, we write $\{\{a_1, \ldots a_n\}\} = \{\{b_1, \ldots, b_n\}\}$. That is, the *multisets* formed by the two sequences are identical.

Definition 5 (Cancellation). *The preference relation $\succeq$ on C satisfies* cancellation *iff for all pairs of sequences $\langle a_1, \ldots, a_n \rangle$ and $\langle b_1, \ldots, b_n \rangle$ of elements of C such that $\{\{a_1, \ldots, a_n\}\} = \{\{b_1, \ldots, b_n\}\}$, if $a_i \succeq b_i$ for $i \leq n-1$, then $b_n \succeq a_n$.*

Roughly speaking, cancellation says that if two collections of choices are identical, then it is impossible to order the choices so as to prefer each choice in the first collection to the corresponding choice in the second collection. The following proposition shows that cancellation is equivalent to reflexivity and transitivity. Although Krantz et al. (1971, p. 251), Fishburn (1987, p. 743) have observed that cancellation implies transitivity, this full characterization appears to be new.

Proposition 1. A preference relation $\succeq$ on a choice set C satisfies cancellation iff

1. $\succeq$ is reflexive, and
2. $\succeq$ is transitive.

Proof. First suppose that cancellation holds. To see that $\succeq$ is reflexive, take $n = 1$ and $a_1 = b_1 = a$ in the cancellation axiom. The hypothesis of the cancellation axiom clearly holds, so we must have $a \succeq a$. To see that cancellation implies transitivity, consider the pair of sequences $\langle a, b, c\rangle$ and $\langle b, c, a\rangle$. Cancellation clearly applies. If $a \succeq b$ and $b \succeq c$, then cancellation implies $a \succeq c$. We defer the proof of the converse to the Appendix. □

We use two strengthenings of cancellation in our representation theorems for $\mathcal{A}$ and $\mathcal{A}^+$, respectively. The first, statewise cancellation, simply increases the set of sequence pairs to which the conclusions of the axiom must apply. This strengthening is required for the existence of additively separable preference representations when choices have a factor structure. Here we state the condition for Savage acts. Given are finite sets S of states and O of outcomes. A *Savage act* is a map $f : S \to O$. Let C denote a set of Savage acts and suppose that $\succeq$ is a preference relation on C.

Definition 6 (Statewise Cancellation)**.** *The preference relation $\succeq$ on a set C of Savage acts satisfies* statewise cancellation *iff for all pairs of sequences $\langle a_1, \ldots, a_n\rangle$ and $\langle b_1, \ldots, b_n\rangle$ of elements of C, if $\{\{a_1(s), \ldots, a_n(s)\}\} = \{\{b_1(s), \ldots, b_n(s)\}\}$ for all $s \in S$, and $a_i \succeq b_i$ for $i \leq n-1$, then $b_n \succeq a_n$.*

Statewise cancellation is a powerful assumption because equality of the multisets is required only 'pointwise'. Any pair of sequences that satisfy the conditions of cancellation also satisfies the conditions of statewise cancellation, but the converse is not true. For instance, suppose that $S = \{s_1, s_2\}$, and we use (o_1, o_2) to refer to an act with outcome o_i in state i, $i = 1, 2$. Consider the two sequences of acts $\langle (o_1, o_1), (o_2, o_2)\rangle$ and $\langle (o_1, o_2), (o_2, o_1)\rangle$. These two sequences satisfy the conditions of statewise cancellation, but not that of cancellation.

In addition to the conditions in Proposition 1, statewise cancellation directly implies *event independence*, a condition at the heart of SEU representation theorems (and which can be used to derive the Sure Thing Principle). If $T \subseteq S$, let $a_T b$ be the Savage act that agrees with a on T and with b on $S - T$; that is $a_T b(s) = a(s)$ if $s \in T$ and $a_T b(s) = b(s)$ if $s \notin T$. We say that $\succeq$ satisfies *event independence* iff for all acts a, b, c, and c' and subsets T of the state space S, if $a_T c \succeq b_T c$, then $a_T c' \succeq b_T c'$.

Proposition 2. If $\succeq$ satisfies statewise cancellation, then $\succeq$ satisfies event independence.

Proof. Take $\langle a_1, a_2 \rangle = \langle a_T c, b_T c' \rangle$ and take $\langle b_1, b_2 \rangle = \langle b_T c, a_T c' \rangle$. Note that for each state $s \in T$, $\{\{a_T c(s), b_T c'(s)\}\} = \{\{a(s), b(s)\}\} = \{\{b_T c(s), a_T c'(s)\}\}$, and for each state $s \notin T$, $\{\{a_T c(s), b_T c'(s)\}\} = \{\{c(s), c'(s)\}\} = \{\{b_T c(s), a_T c'(s)\}\}$. Thus, we can apply statewise cancellation to get that if $a_T c \succeq b_T c$, then $a_T c' \succeq b_T c'$. $\square$

Proposition 1 provides a provides a characterization of cancellation for choices in terms of familiar properties of preferences. We do not have a similarly simple characterization of statewise cancellation. In particular, the following example shows that it is not equivalent to the combination of reflexivity and transitivity of $\succeq$ and event independence.

Example 2. Suppose that $S = \{s_1, s_2\}$, $O = \{o_1, o_2, o_3\}$. There are nine possible acts. Suppose that $\succeq$ is the smallest reflexive, transitive relation such that

$$(o_1, o_1) \succ (o_1, o_2) \succ (o_2, o_1) \succ (o_2, o_2) \succ (o_3, o_1) \succ \\ (o_1, o_3) \succ (o_2, o_3) \succ (o_3, o_2) \succ (o_3, o_3),$$

using the representation of acts described above. To see that $\succeq$ satisfies event independence, note that

- $(x, o_1) \succeq (x, o_2) \succeq (x, o_3)$ for $x \in \{o_1, o_2, o_3\}$;
- $(o_1, y) \succeq (o_2, y) \succeq (o_3, y)$ for $y \in \{o_1, o_2, o_3\}$.

However, statewise cancellation does not hold. Consider the sequences

$$\langle (o_1, o_2), (o_2, o_3), (o_3, o_1) \rangle \text{ and } \langle (o_2, o_1), (o_3, o_2), (o_1, o_3) \rangle.$$

This pair of sequences clearly satisfies the hypothesis of statewise cancellation, that $(o_1, o_2) \succeq (o_2, o_1)$ and $(o_2, o_3) \succeq (o_3, o_2)$, but $(o_1, o_3) \not\succeq (o_3, o_1)$. ∎

For our representation theorems for complete orders, statewise cancellation suffices. However, for partial orders, we need a version of cancellation that is equivalent to statewise cancellation in the presence of **A1**, but is in general stronger.

Definition 7 (Extended Statewise Cancellation)**.** *The preference relation $\succeq$ on a set C of Savage acts satisfies extended statewise cancellation if and only if for all pairs of sequences $\langle a_1, \dots, a_n \rangle$ and $\langle b_1, \dots, b_n \rangle$ of elements of C such that $\{\{a_1(s), \dots, a_n(s)\}\} = \{\{b_1(s), \dots, b_n(s)\}\}$ for all $s \in S$, if there exists some $k < n$ such that $a_i \succeq b_i$ for $i \leq k$, $a_{k+1} = \cdots = a_n$, and $b_{k+1} = \cdots = b_n$, then $b_n \succeq a_n$.*

Proposition 3. In the presence of **A1**, extended statewise cancellation and statewise cancellation are equivalent.

Proof*.* Suppose the hypotheses of extended statewise cancellation hold. If $b_n \succeq a_n$, we are done. If not, by **A1**, $a_n \succeq b_n$. But then the hypotheses of statewise cancellation hold, so again, $b_n \succeq a_n$. $\square$

The extension of cancellation needed for $\mathcal{A}^+$ is based on the same idea as extended statewise cancellation, but probabilities of objects rather than the incidences of objects are added up. Let C denote a collection of elements from a finite-dimensional mixture space. Thus, C can be viewed as a subspace of $\mathbf{R^n}$ for some n, and each component of any $c \in C$ is a probability. We can then formally 'add' elements of C, adding elements of $\mathbf{R^n}$ pointwise. (Note that the result of adding two elements in C is no longer an element of C, and in fact is not even a mixture.)

Definition 8 (Extended Mixture Cancellation)**.** *The preference relation $\succeq$ on C satisfies* extended mixture cancellation *iff for all pairs of sequences $\langle a_1, \dots, a_n \rangle$ and $\langle b_1, \dots, b_n \rangle$ of elements of C, such that $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$, if there exists some $k < n$ such that $a_i \succeq b_i$ for $i \leq k$, $a_{k+1} = \cdots = a_n$, and $b_{k+1} = \cdots = b_n$, then $b_n \succeq a_n$.*

We can extend Proposition 1 to get a characterization theorem for preferences on mixture spaces by using an independence postulate. The preference order $\succeq$ satisfies *mixture independence* if for all a, b, and c in C, and all $r \in (0, 1]$, $a \succeq b$ iff $ra + (1-r)c \succeq rb + (1-r)c$. The preference relation $\succeq$ satisfies *rational mixture independence* if it satisfies mixture independence for all rational $r \in (0, 1]$.

Theorem 3. *A preference relation $\succeq$ on a finite-dimensional mixture space C satisfies extended mixture cancellation iff $\succeq$ is reflexive, transitive, and satisfies rational mixture independence.*

Proof. Suppose that $\succeq$ satisfies extended mixture cancellation. Then it satisfies cancellation, and so from Proposition 1, $\succeq$ is reflexive and transitive. To show that $\succeq$ satisfies rational mixture independence, suppose that $a \succeq b$ and $r = m/n$. Let $a_1 = \cdots = a_m = a$ and $a_{m+1} = \cdots = a_{m+n} = rb + (1-r)c$; let $b_1 = \cdots = b_m = b$ and $b_{m+1} = \cdots = b_{m+n} = ra + (1-r)c$. Then $\sum_{i=1}^{m+n} a_i = \sum_{i=1}^{m+n} b_i$, and so $ra + (1-r)c \succeq rb + (1-r)c$.

Similarly, if $ra + (1-r)c \succeq rb + (1-r)c$, then applying extended mixture cancellation to the same sequence of acts shows that $a \succeq b$.

For the converse, suppose that $\succeq$ is reflexive, transitive, and satisfies rational mixture independence. Suppose that $\langle a_1, \ldots, a_n \rangle$ and $\langle b_1, \ldots, b_n \rangle$ are sequences of of elements of C such that $a_1 + \cdots + a_n = b_1 + \cdots + b_n$, $a_i \succeq b_i$ for $i = 1, \ldots, n-k$, $a_{k+1} = \ldots = a_n$, and $b_{k+1} = \ldots = b_n$. Then from transitivity and rational mixture independence we get that

$$\frac{1}{n}(a_1 + \cdots + a_n) \succeq \frac{1}{n}(b_1 + \cdots + b_k + a_{k+1} + \cdots + a_n)$$
$$= \frac{1}{n}(b_1 + \cdots + b_k) + \frac{n-k}{n} a_n.$$

Since $b_{k+1} = \ldots = b_n$ and $a_1 + \cdots + a_n = b_1 + \cdots + b_n$, we have that

$$\frac{1}{n}(b_1 + \cdots + b_k) + \frac{n-k}{n}(b_n) = \frac{1}{n}(b_1 + \cdots + b_n) = \frac{1}{n}(a_1 + \cdots + a_n).$$

Thus, by transitivity,

$$\frac{1}{n}(b_1 + \cdots + b_k) + \frac{n-k}{n}(b_n) \succeq \frac{1}{n}(b_1 + \cdots + b_k) + \frac{n-k}{n}(a_n).$$

By rational mixture independence, it follows that $b_n \succeq a_n$. $\square$

We can strengthen extended mixture cancellation just as we extended cancellation, by defining a statewise version of it appropriate for AA acts (i.e., functions from S to $\Delta(O)$). For completeness, we give the definition here:

Definition 9 (Extended Statewise Mixture Cancellation)**.** *The preference relation* $\succeq$ *on a set* C *of AA acts satisfies* extended statewise mixture cancellation *iff for all pairs of sequences* $\langle a_1, \ldots, a_n \rangle$ *and* $b = \langle b_1, \ldots, b_n \rangle$ *of elements of* C*, such that* $\sum_{i=1}^{n} a_i(s) = \sum_{i=1}^{n} b_i(s)$ *for all states* s*, if there exists some* $k < n$ *such that* $a_i \succeq b_i$ *for* $i \leq k$*,* $a_{k+1} = \cdots = a_n$*, and* $b_{k+1} = \cdots = b_n$*, then* $b_n \succeq a_n$*.*

It turns out that we do not needed extended statewise mixture cancellation. As the following result shows, it follows from extended mixture cancellation.

Proposition 4. $\succeq$ satisfies extended statewise mixture cancellation iff $\succeq$ satisfies extended mixture cancellation.

Proof*.* Clearly if $\succeq$ satisfies extended statewise mixture cancellation, then it satisfies extended mixture cancellation. For the converse, suppose that $\succeq$ satisfies extended mixture cancellation, $\sum_{i=1}^{n} a_i(s) = \sum_{i=1}^{n} b_i(s)$ for all states s, there exists some $k < n$ such that $a_i \succeq b_i$ $a_{k+1} = \cdots = a_n$, and $b_{k+1} = \cdots = b_n$. Then $(1/n)a_1) + \cdots + (1/n)a_n = (1/n)b_1 + \cdots + (1/n)b_n$. By rational mixture independence (which follows from extended mixture cancellation, by Theorem 3), since $a_1 \succeq b_1$, we have that $(1/n)a_1 + \cdots + (1/n)a_n \succeq (1/n)b_1 + (1/n)a_2 + \cdots + (1/n)a_n$. By a straightforward induction (using transitivity, which again follows from extended mixture cancellation), it follows that

$$\begin{aligned}(1/n)b_1 + \cdots + (1/n)b_k + ((n-k)/n)b_n &= (1/n)a_1 + \cdots + (1/n)a_n \\ &\succeq (1/n)b_1 + \cdots + (1/n)b_k + \\ &\quad ((n-k)/n)a_n.\end{aligned}$$

Now from mixture independence, it follows that $b_n \succeq a_n$, as desired. □

3.3 The cancellation postulate for choices

We use cancellation to get a representation theorem for preference relations on choices. However, the definition of the cancellation postulates for Savage acts and mixtures rely on (Savage) states. We now develop an analogue of this postulate for our syntactic notion of choice.

Definition 10. *Given a set* $T_0 = \{t_1, \ldots, t_n\}$ *of primitive tests, an* atom over T_0 *is a test of the form* $t_1' \wedge \ldots \wedge t_n'$*, where* t_i' *is either* t_i *or* $\neg t_i$*.*

An atom is a possible complete description of the truth value of tests according to the DM. If there are n primitive tests in T_0, there are 2^n atoms. Let $\mathrm{At}(T_0)$ denote the set of atoms over T_0. It is easy to see that, for all tests $t \in T$ and atoms $\delta \in \mathrm{At}(T_0)$, and for all state spaces S and standard test interpretations π_S, either $\pi_S(\delta) \subseteq \pi_S(t)$ or $\pi_S(\delta) \cap \pi_S(t) = \emptyset$. (The formal proof is by induction on the structure of t.) We write $\delta \Rightarrow t$ if the former is the case. We remark for future reference that a standard test interpretation is determined by its behavior on atoms. (It is, of course, also determined completely by its behavior on primitive tests).

Definition 11. *An atom δ (resp., test t) is* consistent with a theory AX *if there exists a state space S and a test interpretation $\pi_S \in \Pi_{AX}$ such that $\pi_S(\delta) \neq \emptyset$ (resp., $\pi_S(t) \neq \emptyset$). Let $\mathrm{At}_{\mathrm{AX}}(T_0)$ denote the set of atoms over T_0 consistent with AX.*

Intuitively, an atom δ is consistent with AX if there is some state in some state space where δ might hold, and similarly for a test t.

A choice in $\mathcal{A}$ can be identified with a function from atoms to primitive choices in an obvious way. For example, if a_1, a_2, and a_3 are primitive choices and $T_0 = \{t_1, t_2\}$, then the choice $a =$ **if** t_1 **then** a_1 **else** (**if** t_2 **then** a_2 **else** a_3) can be identified with the function f_a such that

- $f_a(t_1 \wedge t_2) = f_a(t_1 \wedge \neg t_2) = a_1$;
- $f_a(\neg t_1 \wedge t_2) = a_2$; and
- $f_a(\neg t_1 \wedge \neg t_2) = a_3$.

Formally, we define f_a by induction on the structure of choices. If $a \in \mathcal{A}_0$, then f_a is the constant function a, and

$$f_{\textbf{if } t \textbf{ then } a \textbf{ else } b}(\delta) = \begin{cases} f_a(\delta) & \text{if } \delta \Rightarrow t \\ f_b(\delta) & \text{otherwise.} \end{cases}$$

We consider a family of cancellation postulates, relativized to the axiom system AX. The cancellation postulate for AX (given the language $\mathcal{A}_0$) is simply statewise cancellation for Savage acts, with atoms consistent with AX playing the role of states.

$\mathbf{A2'}$. If $\langle a_1, \ldots, a_n \rangle$ and $\langle b_1, \ldots, b_n \rangle$ are two sequences of choices in $\mathcal{A}_{\mathcal{A}_0,T_0}$ such that for each atom $\delta \in \mathrm{At}_{\mathrm{AX}}(T_0)$, $\{\{f_{a_1}(\delta), \ldots, f_{a_n}(\delta)\}\} = \{\{f_{b_1}(\delta), \ldots, f_{b_n}(\delta)\}\}$, and there exists some $k < n$ such that $a_i \succeq b_i$ for all $i \leq k$, $a_{k+1} = \cdots = a_n$, and $b_{k+1} = \cdots = b_n$, then $b_n \succeq a_n$.

We drop the prime, and refer to $\mathbf{A2}$ when $k = n - 1$.

Axiom $\mathbf{A2}$ implies the simple cancellation of the last section, and so the conclusions of Proposition 1 hold: $\succeq$ on $\mathcal{A}$ will be transitive and reflexive. $\mathbf{A2}$ has another consequence: a DM must be indifferent between AX-equivalent choices.

Proposition 5. Suppose that $\succeq$ satisfies $\mathbf{A2}$. Then $a \equiv_{\mathrm{AX}} b$ implies $a \sim b$.

Proof. Let $S = \mathrm{At}_{\mathrm{AX}}(T_0)$, the set of atoms consistent with AX, let O be $\mathcal{A}_0$, the set of primitive choices, and define $\rho^0_{SO}(c)$ to be the constant function c for a primitive choice c. It is easy to see that $\rho_{SO}(c) = f_c$ for all choices c. If $a \equiv_{\mathrm{AX}} b$, then $\rho_{SO}(a) = \rho_{SO}(b)$, so we must have $f_a = f_b$. Now apply $\mathbf{A2}$ with $a_1 = a$ and $b_1 = b$ to get $b \succeq a$, and then reverse the roles of a and b. $\square$

Proposition 5 implies that the behavior of a and b on atoms not in $\mathrm{At}_{\mathrm{AX}}(T_0)$ is irrelevant; that is, they are null in Savage's sense. We define this formally:

Definition 12. *A test t is* null *if, for all acts a, b and c,* ***if*** *t* ***then*** *a* ***else*** *$c \sim$* ***if*** *t* ***then*** *b* ***else*** *c.*

An atom (or test) inconsistent with the theory AX is null, but consistent tests may be null as well. The notion of a null test is suggestive of, more generally, test-contingent preferences.

Definition 13. *If t is a test in T, then for any acts a and b, $a \succeq_t b$ iff for some c,* ***if*** *t* ***then*** *a* ***else*** *$c \succeq$* ***if*** *t* ***then*** *b* ***else*** *c.*

Proposition 2 shows that statewise cancellation implies that the choice of c is irrelevant, and so test-contingent preferences are well-defined.

To get a representation theorem for $\mathcal{A}^+$, we use a mixture cancellation postulate, again replacing states by atoms. The idea now is that we can identify each choice a with a function f_a mapping atoms consistent with AX into distributions over primitive choices. For example, if t is the only test in T_0 and $\mathrm{AX} = \emptyset$, then the choice $a = \frac{1}{2}a_1 + \frac{1}{2}($**if** t **then** a_2 **else** $a_3)$ can be identified with the function f_a such that

- $f_a(t)(a_1) = 1/2$; $f_a(t)(a_2) = 1/2$
- $f_a(\neg t)(a_1) = 1/2$; $f_a(\neg t)(a_3) = 1/2$.

Formally, we just extend the definition of f_a given in the previous section by defining

$$f_{ra_1+(1-r)a_2}(\delta) = rf_{a_1}(\delta) + (1-r)f_{a_2}(\delta).$$

Consider the following cancellation postulate:

$\mathbf{A2}^\dagger$**.** If $\langle a_1, \ldots, a_n \rangle$ and $\langle b_1, \ldots, b_n \rangle$ are two sequences of acts in $\mathcal{A}^+_{\mathcal{A}_0,T_0}$ such that

$$f_{a_1}(\delta) + \cdots + f_{a_n}(\delta) = f_{b_1}(\delta) + \cdots + f_{b_n}(\delta)$$

for all atoms δ consistent with AX, and there exists $k < n$ such that $a_i \succeq b_i$ for $i \leq k$, $a_{k+1} = \ldots = a_n$, and $b_{k+1} = \ldots = b_n$, then $b_n \succeq a_n$.

Again, $\mathbf{A2}^\dagger$ can be viewed as a generalization of $\mathbf{A2}'$.

$\mathbf{A2}^\dagger$ is analogous to extended *statewise* mixture cancellation. It may seem strange that we need the cancellation to be statewise, since Proposition 4 shows that extended statewise mixture cancellation is equivalent to extended mixture cancellation. This suggests that we might be able to get away with the following simpler axiom:

$\mathbf{A2}^*$**.** If $\langle a_1, \ldots, a_n \rangle$ and $\langle b_1, \ldots, b_n \rangle$ are two sequences of acts in $\mathcal{A}^+_{\mathcal{A}_0,T_0}$ such that

$$f_{a_1} + \cdots + f_{a_n} = f_{b_1} + \cdots + f_{b_n},$$

and there exists $k < n$ such that $a_i \succeq b_i$ for $i = 1, \ldots, k$, $a_{k+1} = \ldots = a_n$, and $b_{k+1} = \ldots = b_n$, then $b_n \succeq a_n$.

$\mathbf{A2}^*$ is implied by $\mathbf{A2}^\dagger$, but they are not quite equivalent. To get equivalence, we need one more property:

EQUIV. If $f_a = f_b$, then $a \sim b$.

EQUIV says that all that matters about a choice is how it acts as a function from tests to primitive choices.

Proposition 6. A preference order $\succeq$ on choices satisfies $\mathbf{A2}^\dagger$ iff it satisfies $\mathbf{A2}^*$ and **EQUIV**.

Proof. If $\succeq$ satisfies $\mathbf{A2}^\dagger$, then clearly $\succeq$ satisfies $\mathbf{A2}^*$. To see that $\succeq$ satisfies **EQUIV**, suppose that $f_a = f_b$. A straightforward argument by induction on the structure of a shows that, for all sets S of states, sets O of outcomes, test interpretations π_S, and choice interpretations ρ_{SO}, we have that $\rho_{SO}(a)(s) = \rho_{SO}(f_a(\delta))$, where δ is the unique atom such that $s \in \pi_S(\delta)$. Similarly, $\rho_{SO}(b)(s) = \rho_{SO}(f_b(\delta))$. Thus, since $f_a = f_b$, $\rho_{SO}(a) = \rho_{SO}(b)$. It follows that $f_a = f_b$ implies $a \equiv b$. It now follows from Proposition 5 that $a \sim b$.

For the converse, since $f_{a_1} + \cdots + f_{a_n} = f_{b_1} + \cdots + f_{b_n}$, it follows that $f_{1/n(a_1)+\cdots+1/n(a_n)} = f_{1/n(b_1)+\cdots+1/n(b_n)}$. Applying **EQUIV**, we conclude that $1/n(a_1) + \cdots + 1/n(a_n) \sim 1/n(b_1) + \cdots + 1/n(b_n)$. We can now continue as in the proof of Proposition 4. $\square$

4 Representation Theorems

Having discussed our framework, we now turn to the representation theorems. Our goal is to be as constructive as possible. In this spirit we want to require that preferences exist not for all possible acts that can be described in a given language, but only for those in a given subset, henceforth designated C. We are agnostic about the source of C. It could be the set of choices in one particular decision problem, or it could be the set of choices that form a universe for a collection of decision problems. One cost of our finite framework is that we will have no uniqueness results. In our framework the preference representation can fail to be unique because of our freedom to choose different state and outcome spaces, but even given these choices, the lack of richness of C may allow multiple representations of the same (partial) order.

4.1 A Representation Theorem for $\mathcal{A}$

By a representation for a preference order on $\mathcal{A}$ we mean the following:

Definition 14. *A preference relation on a set $C \subseteq \mathcal{A}_{\mathcal{A}_0,T_0}$ has a* constructive AX-consistent SEU representation *iff there is a finite set of states S, a finite set O of outcomes, a set $\mathcal{U}$ of utility functions $u : O \rightarrow \mathbf{R}$, a set $\mathcal{P}$ of probability distributions on S, a subset $\mathcal{V} \subseteq \mathcal{U} \times \mathcal{P}$, a test interpretation π_S consistent with AX, and a choice interpretation ρ_{SO} such that $a \succeq b$ iff*

$$\sum_{s\in S} u\big(\rho_{SO}(a)(s)\big)p(s) \geq \sum_{s\in S} u\big(\rho_{SO}(b)(s)\big)p(s) \text{ for all } (u,p) \in \mathcal{V}.$$

We are about to claim that $\succeq$ satisfies $\mathbf{A2}'$ if and only if it has a constructive AX-consistent representation. In the representation, we have a great deal of flexibility as to the choice of the state space S and the outcome space O. One might have thought that the space of atoms, $\mathrm{At}_{\mathrm{AX}}(T_0)$, would be a rich enough state space on which to build representations. This is not true when preferences are incomplete. A rich enough state space needs also to account for the incompleteness.

Definition 15. *Given a partial order $\succeq$ on a set C of choices, let $EX_{\mathrm{AX}}(\succeq)$ denote all the extensions of $\succeq$ to a total order on C satisfying $\mathbf{A2}$.*

Our proof shows that we can take S to be $\mathrm{At}_{\mathrm{AX}}(T_0) \times EX_{\mathrm{AX}}(\succeq)$. Thus, in particular, if $\succeq$ is complete, then we can take the state space to be $\mathrm{At}_{\mathrm{AX}}(T_0)$. We later give examples that show that if $\succeq$ is not complete then, in general, the state space must have cardinality larger than that of $\mathrm{At}_{\mathrm{AX}}(T_0)$. While for some applications there may be a more natural state space, our choice of state space shows that we can always view the DM's uncertainty as stemming from two sources: the truth values of various tests (which are determined by the atom) and the relative order of two choices not determined by $\succeq$ (which is given by the extension $\succeq' \in EX_{\mathrm{AX}}(\succeq)$ of $\succeq$). The idea of a DM being uncertain about her preferences is prevalent elsewhere in decision theory; for instance, in the menu choice literature (Kreps 1979). This uncertainty can be motivated in any number of ways, including both incomplete information and resource-bounded reasoning.

Theorem 4. *A preference relation $\succeq$ on a set $C \subseteq \mathcal{A}_{\mathcal{A}_0,T_0}$ has a constructive AX-consistent SEU representation iff $\succeq$ satisfies $\mathbf{A2}'$. Moreover, in the representation,* *either $\mathcal{P}$ or $\mathcal{U}$ can be taken to be a singleton and, if $\mathcal{U}$ is a singleton $\{u\}$, the state space can be taken to be $\mathrm{At}_{\mathrm{AX}}(T_0) \times EX_{\mathrm{AX}}(\succeq)$. If, in addition, $\succeq$ satisfies **A1**, then $\mathcal{V}$ can be taken to be a singleton (i.e., both $\mathcal{P}$ and $\mathcal{U}$ can be taken to be singletons).*

Theorem 4 is proved in the appendix. The proof proceeds by first establishing a state-dependent representation using $\mathcal{A}_0$ as the outcome space, and then, by changing the outcome space, 'bootstrapping' the representation to an EU representation. This technique shows that, when the state and outcome spaces are part of the representation, there is no difference between the formal requirement for a state-dependent representation and that for a SEU representation. This does not mean that expected utility comes 'for free'; rather, we interpret it to mean that the beliefs/desires formalism that motivates expected utility theory is sensible for the decision problems discussed in this subsection only if the particular outcome space chosen for the representation has some justification external to our theory. We note that if preferences satisfy **A1**, the theorem requires only the cancellation axiom $\mathbf{A2}$ rather than the stronger $\mathbf{A2}'$.

There are no uniqueness requirements on $\mathcal{P}$ or $\mathcal{U}$ in Theorem 4. In part, this is because the state space and outcome space are not uniquely determined. But even if **A1** holds, so that the state space can be taken to be the set of atoms, the probability and the utility are far from unique, as the following example shows.

Example 5. Take $\mathcal{A}_0 = \{a, b\}$, $T_0 = \{t\}$, and $\mathrm{AX} = \emptyset$. Suppose that $\succeq$ is the reflexive transitive closure of the following string of preferences:

$$a \succ \textbf{if } t \textbf{ then } a \textbf{ else } b \succ \textbf{if } t \textbf{ then } b \textbf{ else } a \succ b.$$

Every choice in $\mathcal{A}$ is equivalent to one of these four, so **A1** holds, and we can take the state space to be $S^* = \{t, \neg t\}$. Let $O^* = \{o_1, o_2\}$, and define $\rho^0_{S^*O^*}$ so that $\rho^0_{S^*O^*}(a)$ is the constant function o_1 and $\rho^0_{S^*O^*}(b)$ is the constant function o_2. Now define π_{S^*} in the obvious way, so that $\pi_{S^*}(t) = \{t\}$ and $\pi_{S^*}(\neg t) = \{\neg t\}$. We can represent the preference order by using any probability measure p^* such that $p^*(t) > 1/2$ and any utility function u^* such that $u^*(o_1) > u^*(o_2)$. ∎

As Example 5 shows, the problem really is that the set of actions is not rich enough to determine the probability and utility. By way of contrast, Savage's postulates ensure that the state space is infinite and that there are at least two outcomes. Since the acts are all functions from states to outcomes, there must be uncountably many acts in Savage's framework.

The next example shows that without the completeness axiom **A1**, there may be no representation in which there is only one utility function and the state space is $\mathrm{At}_{\mathrm{AX}}(T_0)$.

Example 6. Suppose that $T_0 = \emptyset$, $\mathrm{AX} = \emptyset$, and $\mathcal{A}_0$ (and hence $\mathcal{A}$) consists of the two primitive choices a and b, which are incomparable. In this case, the smallest state space we can use has cardinality at least 2. For if $|S| = 1$, then there is only one possible probability measure on S, and so a single utility function ranks a and b. Since there is only one atom when there are no primitive propositions, we cannot take the state space to be the set of atoms. There is nothing special about taking $T_0 = \emptyset$ here; similar examples can be constructed showing that we cannot take the state space to be $\mathrm{At}_{\mathrm{AX}}(T_0)$ for arbitrary choices of T_0 if the preference order is partial. An easy argument also shows that there is no representation where $|O| = 1$.

This preference relation can be represented with two outcomes and two states. Let $S = \{s_1, s_2\}$ and $O = \{o_1, o_2\}$. Define $\rho^0_{SO}(a)(s_i) = o_i$, and $\rho^0_{SO}(b)(s_i) = o_{2-i}$ for $i = 1, 2$. Let $\mathcal{U}$ contain a single function such that $u(o_1) \neq u(o_2)$. Let $\mathcal{P}$ be any set of probability measures including the measures p_1 and p_2 such that $p_1(s_1) = 1$ and $p_2(s_1) = 0$. Then the expected utility ranking of randomized acts under each p_i contains no nontrivial indifference, and the ranking under p_2 is the reverse of that under p_1. Thus, these choices represent the preference order. ∎

4.2 A Representation Theorem for $\mathcal{A}^+$

The purpose of this subsection is to show that for the language $\mathcal{A}^+$, we can get something much in the spirit of the standard representation theorem for AA acts. The standard representation theorem has a mixture independence axiom and an Archimedean axiom. As we have seen, $\mathbf{A2}^\dagger$ gives us rational mixture independence; it does not suffice for full mixture independence. To understand what we need, recall that the standard Archimedean axiom for AA acts has the following form:

Arch. If $a \succ b \succ c$ then there exist $r, r' \in (0,1)$ such that $a \succ ra + (1-r)c \succ b \succ r'a + (1-r')c \succ c$.

While this axiom is both necessary for and implied by the existence of an SEU representation when $\succeq$ is complete, the following example describes an incomplete $\succeq$ with a multi-probability SEU representation which fails to satisfy **Arch**.

Example 7. Suppose that $S = \{s_1, s_2, s_3\}$. Let a_1, a_2, and a_3 be acts such that $a_i(s_j)$ gives an outcome of 1 if $i = j$ and 0 otherwise. Let $\mathcal{P}$ consist of all probability distributions p on S such that $p(s_1) \geq p(s_2) \geq p(s_3)$. Define $\succeq$ by taking $a \succeq b$ iff the expected outcome of a is at least as large as that of b with respect to all the probability distributions in $\mathcal{P}$. It is easy to see that $a_1 \succ a_2 \succ a_3$, but for no $r \in (0, 1)$ is it the case that $ra_1 + (1 - r_1)a_3 \succ a_2$ (consider the probability distribution p such that $p(s_1) = p(s_2) = 1/2$). ∎

We can think of the Archimedean axiom as trying to capture some continuity properties of $\succeq$. We use instead the following axiom, which was also used by Aumann (1962, 1964). If the set of tests has cardinality n and the set of primitive choices has cardinality m, we can identify an act $\mathcal{A}^+$ with an element of $\mathbf{R^{2^n m}}$, so the graph of $\succeq$ can be viewed as a subset of $\mathbf{R^{2^{n+1} m}}$.

A3. The graph of the preference relation $\succeq$ is closed.

As the following result shows, in the presence of $\mathbf{A2}^\dagger$ (extended statewise mixture cancellation), $\mathbf{A3}$ implies full mixture independence. Moreover, if we also assume **A1**, then **A3** implies **Arch**. Indeed, it will follow from Theorem 8 that in the presence of **A1** and $\mathbf{A2}^\dagger$, **A3** and **Arch** are equivalent. On the other hand, it seems that **Arch** does not suffice to capture independence if $\succeq$ is a partial order. Summarizing, **A3** captures the essential features of the Archimedean property, while being more appropriate if $\succeq$ is only a partial order.

Proposition 7. (a) $\mathbf{A2}^\dagger$ and $\mathbf{A3}$ imply full mixture independence.
(b) **A1**, $\mathbf{A3}$, and extended mixture cancellation together imply **Arch**.

Proof. For part (a), suppose that $a \succeq b$, and c is an arbitrary act. By Theorem 3, rational mixture independence holds, so we have $ra + (1 - r)c \succeq rb + (1 - r)c$ for all rational r. By $\mathbf{A3}$, we have $ra + (1 - r)c \succeq rb + (1 - r)c$ for all real r. Conversely, suppose that $ra + (1 - r)c \succeq rb + (1 - r)c$ for some real r. If r is rational, it is immediate from rational mixture independence that $a \succeq b$. If r is not rational, choose a rational r' such that $0 < r' < r$. Then we can find a sequence of rational numbers r_n such that $r_n r$ converges to r'. By rational mixture independence, $r_n(ra + (1 - r)c) + (1 - r_n)c \succeq r_n(rb + (1 - r)c) + (1 - r_n)c$. By $\mathbf{A3}$, it follows that $r'a + (1 - r')c \succeq r'b + (1 - r')c$. Now by rational mixture independence, we have $a \succeq b$, as desired.

For part (b), suppose that $a \succ b \succ c$. Mixture independence (which follows from $\mathbf{A2}^\dagger$ and $\mathbf{A3}$, as we have observed) implies that, for all $r \in (0,1)$, $a \succ ra + (1-r)c$. To see that $ra + (1-r)c \succ b$ for some $r \in (0,1)$, suppose not. Then, by **A1**, $b \succeq ra + (1-r)c$ for all $r \in (0,1)$, and by $\mathbf{A3}$, we have that $b \succeq a$, contradicting our initial assumption. The remaining inequalities follow in a similar fashion. $\square$

Definition 16. *A preference relation $\succeq$ on a set $C \subseteq \mathcal{A}^+_{\mathcal{A}_0,T_0}$ has a* constructive AX-consistent SEU representation *iff there is a finite set of states S, a finite set O of outcomes, a set $\mathcal{U}$ of utility functions $u : O \rightarrow \mathbf{R}$, a closed set $\mathcal{P}$ of probability distributions on S, a closed set $\mathcal{V} \subseteq \mathcal{U} \times \mathcal{P}$, a test interpretation π_S consistent with* ***AX****, and a choice interpretation ρ_{SO} such that $a \succeq b$ iff*

$$\sum_{s,o} u(o)\rho_{SO}(a)(s)(o)p(s) \geq \sum_{s,o} u(o)\rho_{SO}(b)(s)(o)p(s) \textit{ for all } (u,p) \in \mathcal{V}.$$

In the statement of the theorem, if $\succeq$ is a preference relation on a mixture-closed subset of $\mathcal{A}^+_{\mathcal{A}_0,T_0}$, we use $\succeq \otimes (a,b)$ to denote the smallest preference order including $\succeq$ and (a,b) satisfying $\mathbf{A2}^\dagger$ and $\mathbf{A3}$, and take $EX^+_{\mathrm{AX}}(\succeq)$ to consist of all complete preference orders extending $\succeq$ and satisfying $\mathbf{A2}^\dagger$ and $\mathbf{A3}$.

Theorem 8. *A preference relation $\succeq$ on a closed and mixture-closed set $C \subseteq \mathcal{A}^+_{\mathcal{A}_0,T_0}$ has a constructive AX-consistent SEU representation iff $\succeq$ satisfies $\mathbf{A2}^\dagger$ and $\mathbf{A3}$. Moreover, in the representation, either $\mathcal{P}$ or $\mathcal{U}$ can be taken to be a singleton and, if $\mathcal{U}$ is a singleton $\{u\}$, the state space can be taken to be $\mathrm{At}_{\mathrm{AX}}(T_0) \times EX^+_{\mathrm{AX}}(\succeq)$. If, in addition, $\succeq$ satisfies* ***A1****, then $\mathcal{V}$ can be taken to be a singleton.*

As in the case of the language $\mathcal{A}$, we cannot in general take the state space to be the set of atoms. Specifically, if $\mathcal{A}_0$ consists of two primitive choices and we take all choices in $\mathcal{A}^+_0$ to be incomparable, then the same argument as in Example 6 shows that we cannot take S to be $\mathrm{At}(T_0)$, and there are no interesting uniqueness requirements that we can place on the set of probability measures or the utility function. On the other hand, if **A1** holds, the proof of Theorem 8 shows that, in the representation, the expected utility is unique up to affine transformations. That is, if $(S, O, p, \pi_S, \rho^0_{SO}, u)$ and $(S', O', p', \pi_{S'}, \rho^0_{S'O'}, u')$ are both representations of $\succeq$, then there exist constants α and β such that for all acts $a \in \mathcal{A}^+_{\mathcal{A}_0,T_0}$, $\mathrm{E}_p(u(\rho_{SO}(a))) = \alpha \mathrm{E}_{p'}(u'(\rho_{S'O'}(a))) + \beta$.

4.3 Objective Outcomes

In choosing, for instance, certain kinds of insurance or financial assets, there is a natural, or objective, outcome space—in these cases, monetary payouts. To model this, we take the set O of objective outcomes as given, and identify it with a subset of the primitive acts A_0. Call the languages with this distinguished set of outcomes $\mathcal{A}_{\mathcal{A}_0,T_0,O}$ and $\mathcal{A}^+_{\mathcal{A}_0,T_0,O}$, depending on whether we allow randomization.

To get a representation theorem in this setting, we need to make some standard assumptions. The first is that there is a best and worst outcome; the second is a state-independence assumption. However, this state-independence assumption only applies to acts in O, but not to all acts.

A4. There are outcomes o_1 and o_0 such that for all non-null tests t, $o_1 \succeq_t a \succeq_t o_0$ for all $a \in \mathcal{A}_0$.

A5. If t is not null and $o, o' \in O$, then $o \succeq o'$ iff $o \succeq_t o'$.

In all our earlier representation theorems, it was possible to use a single utility function. $\mathbf{A4}$ and $\mathbf{A5}$ do not suffice to get such a representation theorem. A necessary condition to have a single utility function, if we also want utility to be state independent, is that $\succeq$ restricted to O be complete.

A6. $\succeq$ restricted to O is complete.

While $\mathbf{A5}$ and $\mathbf{A6}$ are necessary to get a representation with a single utility function, they are not sufficient, as the following example shows.

Example 9. Suppose that we have a language with one primitive test t, and three outcomes, o_0, o_1, and o. Let a_1 be **if** t **then** o_0 **else** o_1 and let a_2 be $\mathbf{if}\ t\ \mathbf{then}\ o_1\ \mathbf{else}\ o_0$. Let $\succeq$ be the smallest preference order satisfying $\mathbf{A2}'$, $\mathbf{A4}$, $\mathbf{A5}$, and $\mathbf{A6}$ (or $\mathbf{A2}^\dagger$, $\mathbf{A3}$, $\mathbf{A4}$, $\mathbf{A5}$, and $\mathbf{A6}$, if we are considering the language $\mathcal{A}^+$) such that $o \sim a_1$ and $o_1 \succ o_0$. Note that a_1 and a_2 are incomparable according to $\succeq$. Suppose that there were a representation of $\succeq$ involving a set $\mathcal{P}$ of probability measures and a single utility function u. Thus, there would have to be probability measures p_1 and p_2 in $\mathcal{P}$ such that $a_1 \succ a_2$ according to (p_1, u) and $a_2 \succ a_1$ according to (p_2, u). It easily follows that $p_1(\pi_S(t)) < 1/2$ and $p_2(\pi_S(t)) > 1/2$. We can assume without loss of generality (by using an appropriate affine transformation) that $u(o_0) = 0$ and

$u(o_1) = 1$. Since $o \sim a_1$, $u(o)$ must be the same as the expected utility of a_1. But this expected utility is less than $1/2$ with p_1 and more than $1/2$ with p_2. This gives the desired contradiction. ∎

Part of the problem is that it is not just the acts in O that must be state independent. Let O^+ be the smallest set of acts containing O that is closed under convex combinations, so that if o and o' are in O^+, then so is $ro + (1-r)o'$. Let $\mathbf{A5}^+$ and $\mathbf{A6}^+$ be the axioms that result by replacing O by O^+ in $\mathbf{A5}$ and $\mathbf{A6}$, respectively. Example 9 actually shows that $\mathbf{A5}^+$ and $\mathbf{A6}$ do not suffice to get a single utility function; Theorem 10 shows that $\mathbf{A5}^+$ and $\mathbf{A6}^+$ do, at least for $\mathcal{A}^+$. We do not have a representation theorem for $\mathcal{A}$, and believe it will be hard to obtain such a theorem (for much the same reasons that it is hard to get a representation theorem in the Savage setting if we restrict to a finite set of acts).

Theorem 10. *A preference relation $\succeq$ on a set $C \subseteq \mathcal{A}^+_{\mathcal{A}_0,T_0,O}$ has a constructive SEU representation with outcome space O iff $\succeq$ satisfies $\mathbf{A2}^\dagger$, $\mathbf{A3}$, $\mathbf{A4}$, $\mathbf{A5}^+$, and $\mathbf{A6}^+$. Moreover, in the representation, $\mathcal{U}$ can be taken to be a singleton $\{u\}$ and the state space can be taken to have the form $\mathrm{At}_{\mathrm{AX}}(T_0) \times \mathcal{A}_0 \times EX^+_{\mathrm{AX}}(\succeq)$. If in addition $\succeq$ satisfies **A1**, then we can take $\mathcal{V}$ to be a singleton too.*

Note that, even if $\succeq$ satisfies $\mathbf{A1}$, the state space has the form $\mathrm{At}_{\mathrm{AX}}(T_0) \times \mathcal{A}_0$. The fact that we cannot take the state space to be $\mathrm{At}_{\mathrm{AX}}(T_0)$ is a consequence of our assumption that primitive acts are deterministic. Roughly speaking, we need the extra information in states to describe our uncertainty regarding how the primitive acts not in O can be viewed as functions from states to outcomes in the pre-specified set O. The following example shows that we need the state space to be larger than $\mathrm{At}_{\mathrm{AX}}(T_0)$ in general.

Example 11. Suppose that there are no primitive proposition, so $\mathrm{At}_{\mathrm{AX}}(T_0)$ is a singleton. There are three primitive acts in $\mathcal{A}_0$: $o_1 = \$50000$, $o_0 = \$0$, and a, which is interpreted as buying 100 shares of Google. Suppose that $o_1 \succ a \succ o_0$. If there were a representation with only one state, then $\rho_{SO}(a)$ would have to be either o_1 or o_0, which would imply that $a \sim o_1$ or $a \sim o_0$, contradicting our description of $\succeq$. The issue here is our requirement that a primitive choice be represented as a function from states to outcomes. If we could represent a as a lottery, there would be no problem representing $\succeq$ with one state. We could simply take $u(o_1) = 1$, $u(o_0) = 0$, and take a to be a lottery that gives each of o_0 and o_1 with probability $1/2$.

We prefer not to allow the DM to consider such ‘subjective’ lotteries. Rather, we have restricted the randomization to acts. The representation theorem would not change if we allowed primitive choice to map a state to a distribution over outcomes, rather than requiring them to be mappings from states to single outcomes (except that we could take the state space to be $\mathrm{At}_{\mathrm{AX}}(T_0)$).

We can easily represent $\succeq$ using two states, s_0 and s_1, by taking a to be the act with outcome o_i in state s_i, for $i = 0, 1$. Taking each of s_0 and s_1 to hold with probability $1/2$ then gives a representation of $\succeq$. In this representation, we can view s_0 as the state where buying Google is a good investment, and s_1 as the state where buying Google is a bad investment. However, the DM cannot talk about Google being a good investment; this is not part of his language. Another DM might explicitly consider the test that Google is a good investment. Suppose this DM has the same preference relation over primitive acts as in the example, is indifferent between 100 shares of Google and o_1 if the test is true, and is indifferent between 100 shares of Google and o_0 if the test is false. This DM's preference order has exactly the same representation as the first DM's preference order, but now s_0 can be viewed as the atom where the test is false, and s_1 can be viewed as the atom where the test is true. The second DM can reason about Google being a good investment explicitly, and can talk to others about it. ∎

5 Nonstandard Hypothesis Interpretations

Many framing problems can be viewed as failures of extensionality, the principle that says that two equivalent descriptions of the same decision problem should lead to the same choices. While this axiom seems appealing, it often fails in laboratory tests. Different descriptions of the same events often give rise to different judgments. Perhaps the best-known example is the medical treatment experiment considered in Example 1, but there are many others. For example, Johnson et al. (1993) found that subjects offered hypothetical health insurance were willing to pay a higher premium for policies covering hospitalization for any disease or accident than they were for policies covering hospitalization for any reason at all.

But what makes two descriptions of events equivalent? In summarizing failures of extensionality, Tversky and Koehler (1994, p. 548) conclude that ‘. . . probability judgments are attached not to events but to descriptions of events.' In our framework, events are described by tests. Whether two sets of tests are recognized by a DM as semantically equivalent is determined by her world view, that is, her theory AX. For instance, suppose that this year's discussion of the failure of the New York Yankees baseball team involves the hypothesis t_1, 'they failed because of a weak defense.' One would have to know more baseball than some of our readers surely do to understand that t_1 is equivalent to $t_2 \vee t_3$, where t_2 is 'they had bad pitching', and t_3 is 'they have poor fielding'. Suppose that a DM must make some decisions to remedy the failures of this year's team. How might her decision problem be represented? If $t_1 \Leftrightarrow t_2 \vee t_3$ is not part of her theory, then the DM's decision problem might be represented by a state space S and a test interpretation π_S such that $\pi_S(t_1) \neq \pi_S(t_2) \cup \pi_S(t_3)$. On the other hand, if her knowledge of baseball is sufficiently refined (as has often been the case for those actually making such decisions), then $t_1 \Leftrightarrow t_2 \vee t_3$ will be part of her theory, and so for any state space S and test interpretation $\pi_s \in \Pi_{AX}$, $\pi_S(t_1) = \pi_S(t_2) \cup \pi_S(t_3)$.

We are not taking a stand on the normative implications of extensionality failures. Our point is that extensionality and its failure is a product of the theory of the world AX that a DM brings to the decision problem at hand (and how it compares to the experimenter's or modeler's view of the world). One can extend conventional expected utility to state spaces and act spaces that admit failures of extensionality. A failure of extensionality is not inconsistent with the existence of an SEU representation on a suitably constructed state space. Moreover, if we modelers take $t_0 \Leftrightarrow t_1$ to be an axiom that describes the world, but our DM disagrees, then the probability of the set $(\pi_S(t_0)/\pi_S(t_1)) \cup (\pi_S(t_1)/\pi_S(t_0))$ measures the degree of framing bias from the modeler's point of view. It could be used in a modeling exercise to qualify the effect of framing failures in particular decision problems, and in empirical studies to measure the degree of framing bias.

Failures of extensionality are concerned with a semantic issue: Identifying when two descriptions of an event are equivalent. However, the same issue arises with respect to logical equivalence. For instance, suppose act a gives a DM x if $\neg((\neg t_1 \wedge \neg t_3) \vee (\neg t_2 \wedge \neg t_3))$ is true, and y otherwise, and act b gives the DM x if $(t_1 \wedge t_2) \vee t_3$ is true, and y otherwise. Only someone adept at formula manipulation (or with a good logic-checking program) will recognize that acts a and b are equivalent as a matter of logic because the equivalence of the two compound propositions is a tautology. So far we have required that DMs be *logically omniscient*, and recognize all such tautologies, because we have considered only standard interpretations.

An interpretation π_S on a state space S does not have to be standard; all that is required is that it associate with each test a subset of S. By allowing nonstandard test interpretations, we can back off from our requirement that DM's know all tautologies. This gives us a way of modeling failures of logical omniscience and, in particular, *resource-bounded reasoning* by DMs.

Recall that a standard test interpretation is completely determined by its behavior on the primitive tests. However, in general, there is no similar finite characterization of a nonstandard test interpretation. To keep things finite, when dealing with nonstandard interpretations, we assume that there is a finite subset T^* of the set T of all tests such that the only tests that appear in choices are those in T^*. (This is one way to model resource-bounded reasoning.) With this constraint, it suffices to consider the behavior of a nonstandard interpretations only on the tests in T^*. Let $\mathcal{A}_{\mathcal{A}_0,T^*}$ consist of all choices whose primitive choices are in $\mathcal{A}_0$ and whose tests are all in T^*.

The restriction to choices in $\mathcal{A}_{\mathcal{A}_0,T^*}$ allows us to define the cancellation postulate in a straightforward way even in the presence of nonstandard interpretations. A *truth assignment to* T^* is just a function $v : T^* \to \{true, false\}$. We can identify an interpretation on S with a function that associates with every state $s \in S$ the truth assignment v_s such that $v_s(t) = true$ iff $s \in \pi(t)$. For a standard interpretation, we can use an atom instead of a truth assignment, since for a standard interpretation, the behavior of each truth assignment is determined by its behavior on primitive propositions, and we can associate with the truth assignment v_s that atom δ_s such that t is a conjunct in δ_s iff $v_s(t) = true$. These observations suggest that we can consider truth assignments the generalization of atoms once we move to nonstandard interpretations. Indeed, if we do this, we can easily generalize all our earlier theorems.

In more detail, we now view a choice as a function, not from atoms to primitive choices, but, more generally, as a function from truth assignments to primitive choices. As before, we take primitive choices to be constant functions. The choice $a =$ **if** t_1 **then** a_1 **else** (**if** t_2 **then** a_2 **else** a_3) can be identified with the function f_a such that

$$f_a(v) = \begin{cases} a_1 & \text{if } v(t_1) = true \\ f_{\textbf{if } t_2 \textbf{ then } a_2 \textbf{ else } a_3} & \text{if } v(t_1) = false \end{cases}$$

and

$$f_{\textbf{if } t_2 \textbf{ then } a_2 \textbf{ else } a_3} = \begin{cases} a_2 & \text{if } v(t_2) = \mathit{true} \\ a_3 & \text{if } v(t_2) = \mathit{false}. \end{cases}$$

A truth assignment v is consistent with AX if $v(t) = \mathit{true}$ for all tests $t \in \mathrm{AX}$.

With these definitions in hand, all our earlier results hold, with the following changes:

- we replace $\mathcal{A}_{\mathcal{A}_0,T_0}$ by $\mathcal{A}_{\mathcal{A}_0,T^*}$;
- we replace 'atoms δ over T_0' by 'truth assignment to T^*'.

The cancellation axioms are all now well defined. With these changes, Proposition 5 and Theorems 4, 8, and 10 hold with essentially no changes in the proof. Thus, we have representation theorems that apply even to resource-bounded reasoners.

6 Updating

There is nothing unique about the state space chosen for an SEU representation of a given choice problem. Our representation theorems state that if an SEU representation exists on any given state space and outcome space with test and choice interpretation functions, then preferences satisfy the appropriate cancellation and other appropriate axioms. Our proofs, however, show that (for standard interpretations) we can always represent a choice situation on the state space $\mathrm{At}_{\mathrm{AX}}(T_0) \times EX_{\mathrm{AX}}(\succeq)$, or $\mathrm{At}_{\mathrm{AX}}(T_0) \times \mathcal{A}_0 \times EX_{\mathrm{AX}}(\succeq)$ for the objective-outcomes case, so this construction is in some sense canonical.

In our models there are two kinds of information. A DM can learn more about the external world, that is, learn the results of some tests. A DM can also learn more about her internal world, that is, she can learn more about her preferences. This learning takes the form of adding more comparisons to her (incomplete) preference order. To make this precise, given a preference order $\succeq$ on a set $C \subseteq \mathcal{A}_{\mathcal{A}_0,T_0}$ satisfying $\mathbf{A2}'$, let $\succeq \oplus (a,b)$ be the smallest preference order including $\succeq$ and (a,b) satisfying $\mathbf{A2}'$. (There is such a smallest preference order, since it is easy to see that if $\succeq'$ and $\succeq''$ both extend $\succeq$, include (a,b), and satisfy $\mathbf{A2}'$, then so does $\succeq' \cap \succeq''$.)

If we take the state space to be $\mathrm{At}_{\mathrm{AX}}(T_0) \times EX_{\mathrm{AX}}(\succeq)$, then a DM's preference order conditioning on either new test information or new comparison information can be represented by conditioning the original probability measures. If $\mathcal{P}$ is a set of probability distributions on some set S and E is a measurable subset of S, let $\mathcal{P} \mid E = \{q : q = p(\cdot|E)$ for some $p \in \mathcal{P}$ with $p(E) > 0\}$. That is, in computing $\mathcal{P} \mid E$, we throw out all distributions p such that $p(E) = 0$, and then apply standard conditioning to the rest. Let $\mathcal{P} \mid t = \mathcal{P} \mid \pi_S(t)$. In the theorems below, we condition on a test t and on a partial order $\succeq'$ extending $\succeq$. We are implicitly identifying t with the event $\{\delta \in \mathrm{At}_{\mathrm{AX}}(T_0) : \delta \Rightarrow t\}$, and $\succeq'$ with the set of total orders in $EX_{\mathrm{AX}}(\succeq)$ extending $\succeq$.

Theorem 4c. Under the assumptions of Theorem 4, and with a representation of $\succeq$ in which $S = \mathrm{At}_{\mathrm{AX}}(T_0) \times EX_{\mathrm{AX}}(\succeq)$ and $\mathcal{U}$ is a singleton $\{u\}$, $\succeq_t$ is represented by $\mathcal{P} \mid t$ and u, and $\succeq \oplus (a, b)$ is represented by $\mathcal{P} \mid (\succeq \oplus (a, b))$ and u.

Theorem 8c. Under the assumptions of Theorem 8, and with a representation of $\succeq$ in which $S = \mathrm{At}_{\mathrm{AX}}(T_0) \times EX^+_{\mathrm{AX}}(\succeq)$ and $\mathcal{U}$ is a singleton $\{u\}$, $\succeq_t$ is represented by $\mathcal{P} \mid t$ and $\{u\}$, and $\succeq \otimes (a, b)$ is represented by $\mathcal{P} \mid (\succeq \otimes (a, b))$ and u.

Theorem 10c. Under the assumptions of Theorem 10, and with a representation of $\succeq$ in which $S = \mathrm{At}_{\mathrm{AX}}(T_0) \times \mathcal{A}_0 \times EX^+_{\mathrm{AX}}(\succeq)$ and $\mathcal{U}$ is a singleton $\{u\}$, $\succeq_t$ is represented by $\mathcal{P} \mid t$ and u, and $\succeq \otimes (a, b)$ is represented by $\mathcal{P} \mid (\succeq \otimes (a, b))$ and u.

Information in the external world is modeled as a restriction on the set of feasible acts; information in the internal world is adding comparisons to a the preference order. These theorems show that both kinds of information can be modeled within a Bayesian paradigm.

7 Conclusion

Our formulation of decision problems has several advantages over more traditional formulations.

- We theorize about only the actual observable choices available to the DM, without having states and outcomes, and without needing to view choices as functions from states to outcomes. Indeed, we show that we can determine whether a DM's behavior is consistent with SEU despite not having states and outcomes. In contrast, in many decision theory experiments, when the DM is given a word problem, the experimenter has an interpretation of this problem as a choice among Savage acts. The experimenter is then really testing whether the DM's choices are consistent with decision theory *given this interpretation*. Thus, a joint hypothesis is being tested. Standard decision theory can be rejected only if the other part of the joint test—that the experimenter and DM interpret represent the word problem with identical Savage acts—is maintained as true. This is unsatisfactory. As we have shown, our approach does place restrictions on choice; thus, our model can be rejected directly by observations of choice between choices observable by the experimenter.

- By viewing choices as *syntactic* objects, our approach allows us to consider DMs that associate different meanings to the same object of choice. Moreover, that meaning can depend on the DM's theory of the world. A DM might have a theory that does not recognize equivalences between tests, and thus choices, that may be obvious to others. This potential difference between a DM's theory of the world and an experimenter's view of the world provides an explanation for framing effects, while still allowing us to view a DM as an expected utility maximizer. Moreover, since a DM's theory may not contain all of standard propositional logic, we can model resource-bounded DMs who cannot discern all the logical consequences of their choices. The existence of an SEU representation and the presence of framing effects are independent issues once one is free to choose a state space. We have modeled framing as a semantic issue, and it appears in the relationship between states of the world and statements about the world in the DMs language. This refutes the unwritten but oft voiced implication that in the presence of extensionality failures, decision theory is irrelevant.

- Our approach allows us to consider different DMs who use different languages to describe the same phenomena. To see why this might be important, consider two decision makers who are interested in 100 shares of Google stock and money (as in Example 11). Suppose that one DM considers quantitative issues like price/earnings ratio to be relevant to the future value of Google, while the other considers astrological tables relevant to Google's future value. The DM who uses astrology might not understand price/earnings ratios (the notion is simply not in his vocabulary) and, similarly, the DM who uses quantitative methods might not understand what it means for the moon to be in the seventh house. Nevertheless, they can trade Google stock and money, as long as they both have available primitive actions like ‘buy 100 shares of Google’ and ‘sell 100 shares of Google’. If we model these decision problems in the Savage framework, we would have to think of assets as Savage acts on a common state and outcome space. Our approach does not require us to tie the DM’s decision problems together with a common state space. Every DM acts as if she has a state space, but these state spaces may be different. Thus, agreeing to disagree results (Aumann 1976), which say that DMs with a common prior must have a common posterior (they cannot agree to disagree) no longer hold.

- A fourth advantage of our approach is more subtle but potentially profound. Representation theorems are just that; they merely provide a description of a preference order in terms of numerical scales. Decision theorists make no pretense that these representations have anything to do with the cognitive processes by which individuals make decisions. But to the extent that the language of choices models the language of the DM, we have the ability to interpret the effects of cognitive limitations having to do with the language in terms of the representation. Our approach allows us to consider the possibility that there may be limitations on the space of choices because some sequence of tests is too computationally costly to verify. Our model of nonstandard test interpretations also takes into account a DM’s potential inability to recognize that two choices logically represent the same function.

There is clearly still much more to do to develop our approach. We briefly mention a few topics here.

- **Learning and dynamic decision-making:** Our focus has been on static decision making. Dealing with decision-making over time will require us to consider learning in more detail. Learning in our framework is not just a matter of conditioning, but also learning about new notions (i.e., becoming aware of new tests). Note that considering dynamic decision-making will require us to take a richer collection of objects of choice, a programming language that allows (among other things) sequential actions (do this, then do that, then do that).

- **Multi-agent decision making:** We have focused on the single-agent case. As we have suggested in examples, once we move to a multi-agent case, we can consider different agents who may use different languages. There is clearly a connection here between our framework and the burgeoning literature on awareness and its applications to game theory (see, for example, (Feinberg 2004, Fagin and Halpern 1988, Halpern 2001, Halpern and Rêgo 2006, Heifetz, Meier, and Schipper 2006, Modica and Rustichini 1999) that needs to be explored.

- **Modeling other 'deviations' from rationality:** We have shown how our approach can model framing problems as a consequence of the agent having a different theory from the modeler. We believe that our approach can also model other 'deviations' from rationality of the type reported by Luce (1990), which can be viewed as a consequence of an agent's bounded processing power. This will require us to be able to distinguish programs such as, for example, $2/3(1/4a + 3/4b) + 1/3c = 1/6a + 5/6(3/5b + 2/5c)$. To do that, we need to give semantics to choices that does not view them as functions from states to distributions over outcomes.

The strategy of 'constructive decision theory' is to use a language to model the DMs statements about the world rather than to insist on Savage's states, construct choices from a given set of primitive objects of choice and propositions in the language, and then to provide semantics for the language and objects of choice in terms of (subjective) states and outcomes. We applied this strategy to choice under uncertainty, but we believe it can be usefully employed in other domains of choice as well, such as intertemporal decision making and interpersonal choice.

Appendix

Proof of Proposition 1. We need to prove the converse part of the proposition. For the converse, suppose that $\succeq$ is reflexive and transitive. By way of contradiction, suppose that $\langle a_1, \ldots, a_n\rangle$ and $\langle b_1, \ldots, b_n\rangle$ are two sequences of minimal cardinality n that violate cancellation; that is, $\{\{a_1, \ldots, a_n\}\} = \{\{b_1, \ldots, b_n\}\}$, $a_i \succeq b_i$ for $i \in \{1, \ldots, n-1\}$, and it is not the case that $b_n \succeq a_n$

If $n = 1$, and $\{\{a\}\} = \{\{b\}\}$, then we must have $a = b$, and the cancellation postulate holds iff $a \succeq a$, which follows from our assumption that $\succeq$ is reflexive.

If $n > 1$, since the two multisets are equal, there must be some permutation τ of $\{1, \ldots, n\}$ such that $a_{\tau(i)} = b_i$. Let $\tau^j(1)$ be the result of applying τ j times, beginning with 1. Let k be the first integer such that $\tau^{k+1}(1)$ is either 1 or n. Then we have the situation described by the following table, where the diagonal arrow denotes equality.

$$
\begin{array}{ccc}
a_1 & \succeq & b_1 \\
 & \swarrow & \\
a_{\tau(1)} & \succeq & b_{\tau(1)} \\
 & \swarrow & \\
\vdots & & \vdots \\
 & \swarrow & \\
a_{\tau^k(1)} & \succeq & b_{\tau^k(1)}
\end{array}
$$

Note that we must have $k \leq n - 1$. If $\tau^{k+1}(1) = 1$, then $b_{\tau^k(i)} = a_1$. Thus, the multisets $\{\{a_1, \ldots, a_{\tau^k(1)}\}\}$ and $\{\{b_1, \ldots, b_{\tau^k(1)}\}\}$ must be equal. The sequences that remain after removing $\{\{a_1, \ldots, a_{\tau^k(1)}\}\}$ from the first sequence and $\{\{b_1, \ldots, b_{\tau^k(1)}\}\}$ from the second also provide a counterexample to the cancellation axiom, contradicting the minimality of n. Thus, $\tau^{k+1}(1) = n$, and we can conclude by transitivity that $a_1 \succeq a_n$.

Continuing on with the iteration procedure starting with $a_{\tau^{k+2}(1)} = a_{\tau(n)} = b_n$, we ultimately must return to a_1 and b_1, as illustrated in the following table: a_1 and b_1.

$$
\begin{array}{ccc}
b_n = a_{\tau^{k+1}(1)} & \succeq & b_{\tau^{k+1}(1)} \\
 & \swarrow & \\
\vdots & & \vdots \\
 & \swarrow & \\
a_{\tau^l(1)} & \succeq & b_{\tau^1(1)} \rightarrow a_1
\end{array}
$$

It follows from transitivity that $b_n \succeq a_1$. By another application of transitivity, we conclude that $b_n \succeq a_n$. This contradicts the hypothesis that the original sequence violated the cancellation axiom. $\square$

We now prove the representation theorems: Theorems 4, 8, and 10. They all use essentially the same technique. It is convenient to start with Theorem 8.

The first step is to get an additively separable utility representation for AA acts on a state space S with outcome space O. This result is somewhat novel because we use extended mixture cancellation and **A3** rather than independence and **Arch**, and because $\succeq$ can be incomplete.

Theorem 12. *A preference relation $\succeq$ on a set C of mixture-closed AA acts mapping a finite set S of states to distributions over a finite set O of outcomes satisfies Extended Mixture Cancellation and **A3** iff there exists a set $\mathcal{U}$ of utility functions on $S \times O$ such that $a \succeq b$ iff*

$$\sum_{s\in S}\sum_{o\in O} u(s,o)a(s)(o) \geq \sum_{s\in S}\sum_{o\in O} u(s,o)b(s)(o) \tag{2}$$

*for all $u \in \mathcal{U}$. Moreover, $\succeq$ also satisfies **A1** iff we can take $\mathcal{U}$ to be a singleton $\{u\}$. In this case, u is unique up affine transformations: if u' also satisfies (2), then there exist α and β such that $u' = \alpha u + \beta$.*

Proof*.* In the totally ordered case, this result is well known. Indeed, for a preference order $\succeq$ that satisfies $\mathbf{A1}$, Proposition 7.4 of (Kreps 1988) shows that such a representation holds iff $\succeq$ satisfies $\mathbf{Arch}$, mixture independence, transitivity, and reflexivity. Theorem 3 and Proposition 7 show that if $\succeq$ satisfies extended cancellation and $\mathbf{A3}$, then these properties hold, so there is a representation. Conversely, if there is such a representation, then all these properties are easily seen to hold. It follows that in the presence of $\mathbf{A1}$, extended cancellation and $\mathbf{A3}$ are equivalent to these properties. However, since we do not want to assume $\mathbf{A1}$, we must work a little harder. Fortunately, the techniques we use will be useful for our later results.

To see that the existence of a representation implies Extended Mixture Cancellation and **A3**, first consider Extended Mixture Cancellation, and suppose that $\langle a_1,\ldots,a_n\rangle$ and $\langle b_1\ldots,b_n\rangle$ are such that $a_1 \succeq b_1$, …, $a_k \succeq b_k$, $a_{k+1} = \ldots = a_n$, $b_{k+1} = \cdots = b_n$, and $a_1+\cdots+a_n = b_1+\cdots+b_n$. For all $u \in \mathcal{U}$, for $i = 1,\ldots,k$, we have

$$\sum_{s\in S} u(s,a_i(s)) \geq \sum_{s\in S} u(s,b_i(s)).$$

Since $a_1+\cdots+a_n = b_1+\cdots+b_n$, for all $s \in S$, it must be that, for all $u \in \mathcal{U}$,

$$\sum_{i=1}^{n}\sum_{s\in S} u(s,a_i(s)) = \sum_{i=1}^{n}\sum_{s\in S} u(s,b_i(s)).$$

Thus, for all $u \in \mathcal{U}$,

$$\sum_{i=k+1}^{n} \sum_{s \in S} u(s, a_i(s)) \leq \sum_{i=k+1}^{n} \sum_{s \in S} u(s, b_i(s)).$$

Since $a_{k+1} = \ldots = a_n$ and $b_{k+1} = \ldots = b_n$, it easily follows that, for all $u \in U$,

$$\sum_{s \in S} u(s, a_n(s)) \leq \sum_{s \in S} u(s, b_n(s)).$$

Thus $b_n \succeq a_n$, as desired. The fact that **A3** holds is straightforward, and left to the reader.

For the 'if' direction, recall that we can view the elements of C as vectors in $\mathbf{R}^{|\mathbf{S}| \times |\mathbf{O}|}$. For the rest of this proof, we identify elements of C with such vectors. Let $D = \{a - b : a, b \in C\}$, and let $D^+ = \{a - b : a \succeq b\}$. Recall that a *(pointed) cone* in $\mathbf{R}^{|\mathbf{S}| \times |\mathbf{O}|}$ is a set CC that is closed under nonnegative linear combinations, so that if $c_1, c_2 \in CC$ and $\alpha, \beta \geq 0$, then $\alpha c_1 + \beta c_2 \in CC$. We need the following lemma.

Lemma 1. *There exists a closed convex cone CC such that $D^+ = CC \cap D$.*

Proof. Let CC consist of all vectors of the form $\alpha_1 d_1 + \cdots + \alpha_n d_n$ for some $n > 0$, where $d_1, \ldots, d_n \in D^+$ and $\alpha_1, \ldots, \alpha_n \geq 0$. Clearly CC is a convex cone, and closed because D^+ is finite. Also, $D^+ \subseteq CC \cap D$. For the opposite inclusion, suppose that $\alpha_1 d_1 + \cdots + \alpha_n d_n = d$, where $d_1, \ldots, d_n \in D^+$, $d \in D$, and $\alpha_1, \ldots, \alpha_n \geq 0$. Thus, there must exist $a_1, \ldots, a_n, b_1, \ldots, b_n, a, b \in C$ such that $a - b = d$, $a_i - b_i = d_i$, and $a_i \succeq b_i$ for $i = 1, \ldots, n$. We want to show that $d \in D^+$ or, equivalently, that $a \succeq b$. Let $r = \alpha_1 + \cdots + \alpha_n + 1$. Since C is mixture-closed, both $(\alpha_1/r)a_1 + \cdots + (\alpha_n/r)a_n + (1/r)b \in C$ and $(\alpha_1/r)b_1 + \cdots + (\alpha_n/r)b_n + (1/r)a$ are in C. Moreover, since $\alpha_1 d_1 + \cdots + \alpha_n d_n = d$ and $a_i \succeq b_i$ for $i = 1, \ldots, n$, it easily follows from mixture independence (which is a consequence of extended mixture cancellation) that $(\alpha_1/r)b_1 + \cdots + (\alpha_n/r)b_n + (1/r)a = (\alpha_1/r)a_1 + \cdots + (\alpha_n/r)a_n + (1/r)b \succeq (1/r)b_1 + \cdots + (1/r)b_n + (1/r)b$. Another application of mixture independence gives us $a \succeq b$, as desired. $\square$

Returning to the proof of Theorem 12, note that if $a \succeq b$ for all a and b, then $D^+ = D$, and we can take CC to be the whole space. If $\succeq$ is nontrivial, then CC is not the whole space. It is well known (Rockafellar 1970) that every closed cone that is not the whole space is the intersection of closed half-spaces (where a half-space is characterized by a vector u such and consists of all the vectors x such that $u \cdot x \geq 0$). Given our identification of elements of C with vectors, we can identify the vector u in $\mathbf{R}^{|\mathbf{S}| \times |\mathbf{O}|}$ characterizing a half-space with a (state-dependent) utility function, where $u(s,o)$ is the (s,o) component of the vector u. If CC is the whole space, we can get a representation by simply taking $\mathcal{U}$ to consist of the single utility function such that $u(s,o) = 0$ for all $(s,o) \in S \times O$. Otherwise, we can take $\mathcal{U}$ to consist of the utility functions characterizing the half-spaces containing CC. It is easy to see that for $a, b \in C$, we have that $a \succeq b$ iff $a - b \in D^+$ iff $a - b \in CC$ iff $u \cdot (a-b) \geq 0$ for every half-space u containing CC; i.e. iff (2) holds. □

To prove Theorem 8, the following lemma, which shows that we can identify complete preference orders with half-spaces, is also useful. Given a subset R of $\mathbf{R}^{|\mathbf{S}| \times |\mathbf{O}|}$, define the relation $\succeq_R$ on C by taking $a \succeq_R b$ iff $a - b \in R$.

Lemma 2. $EX_{\mathrm{AX}} = \{\succeq_R$: R *is either a half-space containing* CC *or the full space*$\}$.

Proof. If R is the full space, then $\succeq_R$ is the trivial relation, so clearly $\succeq_R \in EX_{\mathrm{AX}}(\succeq)$. If R is a half-space H containing CC and H is characterized by u, then $\succeq_H$ extends $\succeq$, since $CC \subseteq H$. To see that $\succeq_H$ satisfies $\mathbf{A1}$, observe that if $(a,b) \notin \succeq_H$, then $u \cdot (a-b) < 0$, so $u \cdot (b-a) > 0$, and $b \succeq_H a$. To see that $\succeq_H$ satisfies $\mathbf{A2'}$, suppose that $a_1 + \cdot + a_n = b_1 + \cdot + b_n$, and $a_i \succeq_H b_i$ for $i = 1, \ldots, n-1$. Thus, $u \cdot (a_1 + \cdot + a_n) = u \cdot (b_1 + \cdot b_n)$, and $u \cdot (a_i - b_i) \geq 0$ for $i = 1, \ldots, n-1$. It follows that $(b_n - a_n) \cdot u \geq 0$, so $b_n \succeq_H a_n$. Finally, for **A3**, it is clear that if $(a_n, b_n) \rightarrow (a,b)$, and $u \cdot (a_n - b_n) \geq 0$, then $u \cdot (a-b) \geq 0$, so $a \succeq_H b$. Thus, $\succeq_H \in EX_{\mathrm{AX}}(\succeq)$.

For the opposite inclusion, suppose that $\succeq' \in EX_{\mathrm{AX}}(\succeq)$. Let CC' be the cone determined by $\succeq'$, as in Lemma 1. Clearly $CC \subseteq CC'$. If CC' is the full space, then we are done, since $\succeq' = \succeq_{CC'}$. Otherwise, CC' is the intersection of half-spaces. Choose a half-space H such that $CC' \subseteq H$. We claim that $\succeq' = \succeq_H$. Suppose not. Since $CC' \subseteq H$, we must have $\succeq' \subseteq \succeq_H$. There must exist $a, b \in C$ such that $a \succeq_H b$ and $a \not\succeq' b$. Since $\succeq'$ is complete, we must have $b \succ' a$. Thus, $b \succeq_H a$, so $a \sim_H b$. Since H is not the full space, there must be some c such that $b \not\sim_H c$. Suppose that $c \succ_H b$. We must have $c \succ b$, since otherwise $b \succeq' c$, and it follows that $b \sim_H c$. By the Archimedean property (which holds by Proposition 7), since $c \succ' b \succ' a$, there exists $r > 0$ such that $b \succ' rc + (1-r)a$. Thus we must have $b \succeq_H rc + (1-r)a \succeq_H b$. But this contradicts the assumption that $c \succ_H b \sim_H a$. We get a similar contradiction if $b \succ_H c$, since then $a \succ_H c$. □

Proof of Theorems 8 and 8c. It is easy to check that if there is a constructive SEU representation of $\succeq$, then $\succeq$ satisfies $\mathbf{A2}^{\dagger}$ and $\mathbf{A3}$.

For the converse, suppose that $\succeq$ satisfies $\mathbf{A2}^{\dagger}$ and $\mathbf{A3}$. Take $S = \mathrm{At}_{\mathrm{AX}}(T_0)$ and $O' = \mathcal{A}_0$. Define $\pi_S(t)$ to be the set of all atoms δ in $\mathrm{At}_{\mathrm{AX}}(T_0)$ such that $\delta \Rightarrow t$. Define $\rho^0_{SO'}(a)$ to be the constant function a for a primitive choice a. It is easy to see that $\rho_{SO'}(a) = f_a$ for all choices $a \in C$. Define a preference relation $\succeq_S$ on the AA acts of the form $\rho_{SO'}(a)$ by taking $f_a \succeq_S f_b$ iff $a \succeq b$. The fact that $\succeq_S$ is well defined follows from Proposition 5, for if $f_a = f_{a'}$, then it easily follows that $a \equiv_{\mathrm{AX}} a'$, so $a \sim a'$. Clearly $\succeq_S$ satisfies extended statewise cancellation, and satisfies **A1** iff $\succeq$ does. Thus, by Theorem 12, there is an additively separable representation of $\succeq$.

Now we adjust the state and outcome spaces to get a constructive SEU representation. Suppose first that **A1** holds. Take $S = \mathrm{At}_{\mathrm{AX}}(T_0)$ and take $O = \mathrm{At}_{\mathrm{AX}}(T_0) \times \mathcal{A}_0$. For a primitive choice $a \in \mathcal{A}_0$, define $\rho^0_{SO}(a)(\delta) = (\delta, a)$. To complete the proof, it clearly suffices to find a probability measure p on $\mathrm{At}_{\mathrm{AX}}(T_0)$ and a utility function v on $\mathrm{At}_{\mathrm{AX}}(T_0) \times \mathcal{A}_0$ such that $u(\delta, a) = p(\delta)v(\delta, a)$, where u is the state-dependent utility function whose existence is guaranteed by Theorem 12. This is accomplished by taking *any* probability measure p on $\mathrm{At}_{\mathrm{AX}}(T_0)$ such that for all atoms δ, $p(\delta) > 0$, and taking $v(\delta, a) = u(\delta, a)/p(\delta)$.

If **A1** does not hold, then proceed as above, using Theorem 12 to get an entire set $\mathcal{U}'$ of functions $u : \mathrm{At}_{\mathrm{AX}}(T_0) \times \mathcal{A}_0 \rightarrow \mathbf{R}$, and a single probability distribution p that assigns positive probability to every atom. Let $\mathcal{U}$ consist of all utility functions u such that there exists some $u' \in \mathcal{U}'$ such that $u(\alpha, \delta) = u'(\delta, a)/p(\delta)$. In this representation, again, the state space is $\mathrm{At}_{\mathrm{AX}}(T_0)$.

We now give a representation using a single utility function. Let $S' = \mathrm{At}_{\mathrm{AX}}(T_0) \times EX_{\mathrm{AX}}(\succeq)$, and let $O'' = \mathrm{At}_{\mathrm{AX}}(T_0) \times \mathcal{A}_0 \times EX_{\mathrm{AX}}(\succeq)$. Define $\rho^0_{S'O''}(a)(\delta, \succeq') = (\delta, a, \succeq')$. For $t \in T_0$, define $\pi_{S'}(t) = \pi_S(t) \times EX_{\mathrm{AX}}(\succeq)$. As before, let $\mathcal{U}'$ be the set of utility functions on $\mathrm{At}_{\mathrm{AX}}(T_0) \times \mathcal{A}_0$ that represent $\succeq$. Lemma 2 shows that $\mathcal{U}'$ consists of one utility function $u_{\succeq'}$ for every total order $\succeq' \in EX_{\mathrm{AX}}(\succeq)$. Again, fix a probability measure p on $\mathrm{At}_{\mathrm{AX}}(T_0)$ such that $p(\delta) > 0$ for all $\delta \in \mathrm{At}_{\mathrm{AX}}(T_0)$. For each relation $\succeq' \in EX_{\mathrm{AX}}(\succeq)$, define $p_{\succeq'}$ on $\mathrm{At}_{\mathrm{AX}}(T_0) \times EX_{\mathrm{AX}}(\succeq)$ by taking $p_{\succeq'}(\delta, \succeq'') = p(\delta)$ if $\succeq' = \succeq''$, and $p_{\succeq'}(\delta, \succeq'') = 0$ if $\succeq' \neq \succeq''$. Let $\mathcal{P} = \{p_{\succeq'} : \succeq' \in EX_{\mathrm{AX}}(\succeq)\}$. Define $v(\delta, a, \succeq') = u_{\succeq'}(\delta, a)/p(\delta)$. It is easy to see that $\mathcal{P} \times \{v\}$ represents $\succeq$. Moreover, it easily follows that, with this representation, $\succeq_t$ and $\succeq \oplus (a, b)$ can be represented by updating. $\square$

Now we show how the ideas in this proof can be modified to prove Theorems 4 and 4c.

Proof of Theorems 4 and 4c. Again, it is easy to check that if there is a constructive SEU representation of $\succeq$, then $\succeq$ satisfies $\mathbf{A2}'$.

For the converse, given a preference relation $\succeq$ that satisfies $\mathbf{A2}'$. The structure of the proof is identical to that of Theorem 8. We first prove an analogue of Theorem 12.

Theorem 13. *A preference relation $\succeq$ on a set C of Savage acts mapping a finite set S of states to a finite set O of outcomes satisfies extended statewise cancellation iff there exists a set $\mathcal{U}$ of utility functions on $S \times O$ such that $a \succeq b$ iff*

$$\sum_{ss \in S} u\big(s, a(s)\big) \geq u\big(s, b(s)\big) \text{ for all } u \in \mathcal{U}. \tag{3}$$

*Moreover, $\succeq$ satisfies **A1** iff $\mathcal{U}$ can be chosen to be a singleton.*

The proof of Theorem 13 is identical to that of Theorem 12, except that we need an analogue of Lemma 1 for the case that $\succeq$ satisfies $\mathbf{A2}'$. Let D and D^+ be defined as in Lemma 1.

Lemma 3. *If $\succeq$ satisfies $\mathbf{A2}'$ There exists a cone CC such that $D^+ = CC \cap D$.*

Proof. Again, let CC consist of all vectors of the form $\alpha_1 d_1 + \cdots + \alpha_n d_n$ for some $n > 0$, where $d_1, \ldots, d_n \in D^+$ and $\alpha_1, \ldots, \alpha_n \geq 0$. Clearly CC is a cone. Since C is closed and bounded, so is D, and **A3** implies that D^+ is closed and bounded. Therefore CC is closed. Furthermore, $D^+ \subseteq CC \cap D$. For the converse inclusion, suppose that $\alpha_1 d_1 + \cdots + \alpha_n d_n = d$, where $d_1, \ldots, d_n \in D^+$, $d \in D$, and $\alpha_1, \ldots, \alpha_n \geq 0$. That means that $\alpha_1, \ldots, \alpha_n$ is a nonnegative solution to the system of equations $x_1 d_1 + \cdots + x_n d_n = d$. Since all the coefficients in these equations are rational (in fact, they are all 0, 1, and -1) there exists a nonnegative rational solution to this system of equations. It easily follows that there exist positive integers $\beta_1, \ldots, \beta_{n+1}$ such that $\beta_1 d_1 + \cdots + \beta_n d_n = \beta_{n+1} d$. By definition, there must exist $a_1, \ldots, a_n, b_1, \ldots, b_n, a, b \in C$ such that $a - b = d$, $a_i - b_i = d_i$, and $a_i \succeq b_i$ for $i = 1, \ldots, n$. It follows that $\{\{a_1, \ldots, a_1, \ldots, a_n, \ldots, a_n, b, \ldots, b\}\} = \{\{b_1, \ldots, b_1, \ldots, a_n, \ldots, a_n, b, \ldots, b\}\}$, where a_i occurs in the left-hand multiset β_i times and b occurs β_{n+1} times, and, similarly, b_i occurs in the right-hand side β_i times and a occurs β_{n+1} times. $\mathbf{A2}'$ now implies that $a \succeq b$, so $d \in CC$, as desired. $\square$

The proof of Theorem 13 is now identical to that of Theorem 12. Moreover, the proof of Theorem 4 now follows from Theorem 13 in exactly the same way that the proof of Theorem 8 follows from Theorem 12. $\square$

Proof of Theorems 10 and 10c. First, let $S = \mathrm{At}_{\mathrm{AX}}(T_0)$ and $O' = \mathcal{A}_0$. As in the proof of Theorem 8, using Theorem 12, we can find an additively separable representation of $\succeq$; that is, we can find a set $\mathcal{U}$ of utility functions on $S = \mathrm{At}_{\mathrm{AX}}(T_0) \times \mathcal{A}_0$ that represent $\succeq$. Let o_1 and o_0 denote the best and worst outcomes guaranteed to exist by $\mathbf{A4}$. Note that it follows from $\mathbf{A4}$ and $\mathbf{A5}$ that $u(\delta, o_0) \leq u(\delta, a) \leq u(\delta, o_1)$ for all $(\delta, a) \in S$. (For null atoms δ, we must in fact have $u(\delta, o_0) = u(\delta, a) = u(\delta, o_1)$ for all $a \in \mathcal{A}_0$). Furthermore, note that we can replace u by u', where $u'(\delta, a) = u(\delta, a) - u(\delta, o_0)$ for all $(\delta, a) \in S$ to get an equivalent representation; thus, by appropriate scaling, we can assume without loss of generality that, for all $u \in \mathcal{U}$, we have that $u(\delta, o_0) = 0$ for all $\delta \in \mathrm{At}_{\mathrm{AX}}(T_0)$ (so $u(\delta, a) \geq 0$ for all $(\delta, a) \in S$) and that $\sum_{\delta' \in \mathrm{At}_{\mathrm{AX}}(T_0)} u_{\succeq'}(\delta', o_1) = 1$. Finally, note that it easily follows from $\mathbf{A3}$ that, for all $o \in O$, there exists a unique $c_o \in [0, 1]$ such that $o \sim c_o o_1 + (1 - c_o) o_0$ (in fact, $c_o = \inf\{c : co_1 + (1-c)o_0 \succeq o\}$). Clearly $c_{o_1} = 1$ and $c_{o_0} = 0$. By $\mathbf{A5}$, it follows that $o \sim_\delta c_o o_1 + (1 - c_o) o_0$ for all atoms δ. Thus, we must have that $u(\delta, o) = c_0 u(\delta, o_1)$ for all atoms δ and all $u \in \mathcal{U}$.

We now construct a state-independent SEU representation using O as the outcome space. Let $S' = \mathrm{At}_{\mathrm{AX}}(T_0) \times \mathcal{A}_0 \times EX^+_{\mathrm{AX}}(\succeq)$. Define $\pi^0_{S'}$ by taking $\pi^0_S(t) = \cup_{\delta \Rightarrow t}\{\delta \times A_0 \times EX^+_{\mathrm{AX}}(\succeq)\}$, and define $\rho_{SO}(a)((\delta, a', \succeq'))$ to be a if $a \in O$; o_1 if $a \in A_0 - O$ and $a \succeq'_\delta a'$; and o_0 otherwise. Let u' be defined by taking $u'(o) = c_o$. Finally, recall that we can take $\mathcal{U} = \{u_{\succeq'} : \succeq' \in EX^+_{\mathrm{AX}}(\succeq)\}$, where $u_{\succeq'}$ represents $\succeq'$. Let $p_{\succeq'}$ be defined so that $p_{\succeq'}(\delta, a, \succeq'') = 0$ and, for all $a \in \mathcal{A}_0$, $p_{\succeq'}(\{(\delta, a', \succeq') : a \succeq' a'\}) = u_{\succeq'}(\delta, a)$. It is easy to check that a probability measure $p_{\succeq'}$ can be defined so as to satisfy this constraint. In particular, note that $p_{\succeq'}(\{\delta\} \times \mathcal{A}_0 \times \{\succeq'\}) = u_{\succeq'}(\delta, o_1)$. For all $(\delta, a) \in S$, we have that

$$u_{\succeq'}(\delta, a) = \sum_{a' :\, a \succeq' a'} p_{\succeq'}(\delta, a', \succeq') = \sum_{a' \in \mathcal{A}_0} p_{\succeq'}(\delta, a', \succeq') u(\rho_{S'O}(a)(\delta, a', \succeq')).$$

It follows that $\mathcal{P}$ and u represent $\succeq$, where $\mathcal{P} = \{p_{\succeq'} : \succeq' \in EX^+_{\mathrm{AX}}(\succeq)\}$. As usual, it is straightforward to verify that updating works appropriately. $\square$

References

AHN, D. (2008): "Ambiguity without a state space," *Review of Economic Studies*, 75(1), 3–28.

AHN, D., AND H. ERGIN (2007): "Framing contingencies," unpublished manuscript.

ANSCOMBE, F. J., AND R. J. AUMANN (1963): "A definition of subjective utility," *Annals of Mathematical Statistics*, 34, 199–205.

AUMANN, R. J. (1962): "Utility theory without the completeness axiom," *Econometrica*, 30, 445–462.

——— (1964): "Utility theory without the completeness axiom: A correction," *Econometrica*, 32, 210–212.

——— (1976): "Agreeing to disagree," *Annals of Statistics*, 4(6), 1236–1239.

BEWLEY, T. (2002): "Knightian Decision Theory," *Decisions in Economics and Finance*, 25(79–110).

BLUME, L., D. EASLEY, AND J. HALPERN (2006): "Redoing the Foundations of Decision Theory," in *Principles of Knowledge Representation and Reasoning: Proc. Tenth International Conference (KR '06)*, pp. 14–24.

DEKEL, E., B. LIPMAN, AND A. RUSTICHINI (2001): "Representing preferences with a unique subjective state space," *Econometrica*, 69, 891–934.

ELIAZ, K., AND E. OK (2006): "Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences," *Games and Economic Behavior*, 56, 61–86.

ELLSBERG, D. (2001): *Risk, Ambiguity and Decision*. Garland Publishing, Inc., New York.

FAGIN, R., AND J. Y. HALPERN (1988): "Belief, Awareness, and Limited Reasoning," *Artificial Intelligence*, 34, 39–76.

FEINBERG, Y. (2004): "Subjective reasoning—games with unawareness," Discussion Paper Resarch Paper Series #1875, Stanford Graduate School of Business.

FISHBURN, P. C. (1987): "Reconsiderations in the foundations of decision under uncertainty," *The Economic Journal*, 97(388), 825–841.

GHIRARDATO, P. (2001): "Coping with ignorance: unforeseen contingencies and non-additive uncertainty," *Economic Theory*, 17, 247–276.

GILBOA, I., AND D. SCHMEIDLER (2004): "Subjective distributions," *Theory and Decision*, 56, 345–357.

HALPERN, J. Y. (2001): "Alternative semantics for unawareness," *Games and Economic Behavior*, 37, 321–339.

HALPERN, J. Y., AND L. C. RÊGO (2006): "Extensive Games with Possibly Unaware Players," in *Proc. Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 744–751. Full version available at arxiv.org/abs/0704.2014.

HEIFETZ, A., M. MEIER, AND B. SCHIPPER (2006): "Interactive unawareness," *Journal of Economic Theory*, 130, 78–94.

JOHNSON, E., J. HERSHEY, J. MESZAROS, AND H. KUNREUTHER (1993): "Framing, probability distortions, and insurance decisions," *Journal of Risk and Uncertainty*, 7, 35–51.

KARNI, E. (2006): "Subjective expected utility theory without states of the world," *Journal of Mathematical Economics*, 42, 325–342.

KRANTZ, D. H., R. D. LUCE, P. SUPPES, AND A. TVERSKY (1971): *Foundations of Measurement, Vol 1: Additive and Polynomial Representations*. Academic Press, New York.

KREPS, D. (1992): "Static choice and unforeseen contingencies," in *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn*, ed. by P. Dasgupta, D. Gale, and O. Hart. MIT Press, Cambridge, Mass.

KREPS, D. M. (1979): "A representation theorem for "preference for flexibility"," *Econometrica*, 47(3), 565–77.

KREPS, D. M. (1988): *Notes on the Theory of Choice*. Westview Press, Boulder, Colo.

LIPMAN, B. (1999): "Decision theory without logical omniscience: Toward an axiomatic framework for bounded rationality," *Review of Economic Studies*, 66, 339–361.

LUCE, R. D. (1990): “Rational versus plausible accounting equivalences in preference judgments,” *Psychological Science*, 1, 225–234, Reprinted with minor changes in Ward Edwards (Ed.), *Utility Theories: Measurements and Applications*, pp. 187–206. Boston: Kluwer, 1992.

MACHINA, M. (2006): “States of the World and State of Decision Theory,” in *The Economics of Risk*, ed. by D. Meyer. W. E. Upjohn Institute, Kalamazoo, MI.

MANDLER, M. (2001): “A difficult choice in preference theory: rationality implies completeness or transitivity but not both,” in *Varieties of Practical Reasoning*, ed. by E. Millgram. MIT Press, Cambridge MA.

MCNEIL, B. J., S. J. PAUKER, H. C. SOX JR., AND A. TVERSKY (1982): “On the elicitation of preferences for alternative therapies,” *New England Journal of Medicine*, 306, 1259–1262.

MODICA, S., AND A. RUSTICHINI (1999): “Unawareness and partitional information structures,” *Games and Economic Behavior*, 27(2), 265–298.

RANTALA, V. (1982): “Impossible worlds semantics and logical omniscience,” *Acta Philosophica Fennica*, 35, 18–24.

ROCKAFELLAR, R. T. (1970): *Convex Analysis*. Princeton University Press, Princeton, N.J.

SAVAGE, L. J. (1954): *Foundations of Statistics*. Wiley, New York.

SCHMEIDLER, D. (1989): “Subjective probability and expected utility without additivity,” *Econometrica*, 57, 571–587.

SCOTT, D. (1964): “Measurement structures and linear inequalities,” *Journal of Mathematical Psychology*, 1, 233–247.

TVERSKY, A., AND D. KAHNEMAN (1992): “Advances in prospect theory: cumulative representation of uncertainty,” *Journal of Risk and Uncertainty*, 5, 297–323.

TVERSKY, A., AND D. J. KOEHLER (1994): “Support theory: A nonextensional representation of subjective probability,” *Psychological Review*, 101(4), 547–567.

WAKKER, P., AND A. TVERSKY (1993): “An axiomatization of cumulative prospect theory,” *Journal of Risk and Uncertainty*, 7(7), 147–176.

WEIBULL, J. (2004): “Testing Game Theory,” in *Advances in Understanding Strategic Behaviour: Game Theory, Experiments and Bounded Rationality. Essays in Honour of Werner Güth*, ed. by S. Huck, pp. 85–104. Palgrave MacMillan.