Theoretical Economics 10 (2015), 489–541 1555-7561/20150489

Hidden actions and preferences for timing of resolution of uncertainty

Haluk Ergin
Department of Economics, University of California–Berkeley

Todd Sarver
Department of Economics, Duke University

We study preferences for timing of resolution of objective uncertainty in a menu-choice model with two stages of information arrival. We characterize a general class of utility representations called hidden action representations, which interpret an intrinsic preference for timing of resolution of uncertainty as if an unobservable action is taken between the resolution of the two periods of information arrival. These representations permit a richer class of preferences for timing than was possible in the model of Kreps and Porteus (1978) by incorporating a preference for flexibility. Our model contains several special cases where this hidden action can be given a novel economic interpretation.

Keywords. Temporal preferences, preference for flexibility, hidden action, subjective uncertainty.

JEL classification. D81.

1. Introduction

This paper considers several new classes of dynamic preferences, providing a utility representation for preferences for early resolution of uncertainty. The first purpose of this analysis is to unite two strands of the literature: We consider a model in which an individual may have an intrinsic preference for timing of resolution of uncertainty (as in Kreps and Porteus 1978) while at the same time exhibiting a preference for flexibility (as in Kreps 1979 and Dekel, Lipman, and Rustichini 2001, henceforth DLR). As we discuss later in the Introduction, preferences exhibiting this combination are quite plausible in a variety of economic environments and can have important implications. The second purpose of this analysis is to provide a simple and intuitive interpretation for such preferences: We provide a representation that suggests that intrinsic preferences for timing of resolution of uncertainty can be interpreted as being the result of some interim action

Haluk Ergin: hie@econ.berkeley.edu
Todd Sarver: todd.sarver@duke.edu
We thank the National Science Foundation (Grants SES-0850980 and SES-0851066) for financial support, and David Ahn, Eddie Dekel, Pablo Schenone, Marciano Siniscalchi, and the editor and two anonymous referees for helpful comments. These results were previously distributed under the title "A Subjective Model of Temporal Preferences."

DOI: 10.3982/TE1340

that is not observable to the modeler (a *hidden* action). Thus, intrinsic preference for timing can be understood as an extrinsic (or instrumental) preference for timing arising due to some unobserved action.

1.1 *Intrinsic versus extrinsic preferences for timing*

It is well known that an individual may prefer to have uncertainty resolve at an earlier date so as to be able to condition her future actions on the realization of this uncertainty. For example, an individual may prefer to have uncertainty about her future income resolve earlier so that she can optimally smooth her consumption across time. Suppose an individual has the possibility of receiving a promotion with a substantial salary increase several years into the future. If she is able to learn the outcome of that promotion decision now, then even if she will not actually receive the increased income until a later date, she may choose to increase her current consumption by temporarily decreasing her savings or increasing her debt. On the other hand, if she is not told the outcome of the promotion decision, then by increasing her consumption now, she risks having larger debt and, hence, suboptimally low consumption in the future. In this example, changing the timing of the resolution of uncertainty benefits the individual by increasing her ability to condition her choices on the outcome of that uncertainty.

Kreps and Porteus (1978) considered a model that enriches the standard additive dynamic expected-utility model by allowing for a preference for early resolution of uncertainty even when the individual's ability to condition her (observed) actions on the outcome of this uncertainty does not change with the timing of its resolution. For example, suppose the individual described above has no current savings and is unable to take on debt. Then if she learns the outcome of the promotion decision now, she is unable to increase her current consumption. Even in this case, the preferences considered by Kreps and Porteus (1978) allow the individual to have a strict preference for that uncertainty to resolve earlier, which we refer to as an intrinsic preference for the timing of the resolution of uncertainty. The additional flexibility of their model has proven useful in applications to macroeconomic models of asset pricing (Epstein and Zin 1989, 1991), precautionary savings (Weil 1993), and business cycles (Tallarini 2000) (see Backus et al. 2004 for a survey of these and related papers).

While an intrinsic preference for early resolution of uncertainty occurs by definition in the absence of any directly observable payoff-relevant action, it is possible that the individual does, in fact, take a payoff-relevant action that is simply unobservable to the economic modeler. For example, suppose the individual described above is not permitted to save or borrow, yet still exhibits a preference for early resolution of uncertainty about future income. It may be the case that this individual has some additional unobserved payoff-relevant action that she would like to condition on the resolution of this uncertainty. This could be a physical action that happens to be unobserved by the modeler, such as choosing durable consumption goods today that better complement high or low future consumption, or it could be psychological in nature, such as some form of mental preparation for the future. In either case, her apparent intrinsic preference for early resolution of uncertainty could, in fact, be an extrinsic preference arising due to an unobserved action.

Kreps and Porteus (1979) provided an interpretation along these lines for the preferences considered in their 1978 paper, and Machina (1984) considered a related representation for slightly more general preferences. Our main representation theorem provides a similar *hidden action* interpretation for a broader class of preferences. This generalization not only allows us to model some useful preferences that have not been previously considered, but also permits novel psychological interpretations for the hidden action, such as costly decision-making.

1.2 *Overview of results*

We examine dynamic preferences in a simple menu-choice setting with two stages of objective uncertainty. This framework is a two-stage version of the environment considered by Kreps and Porteus (1978). However, we allow for more general axioms, which permits us to model a richer set of preferences for early or late resolution of uncertainty. In particular, we incorporate a preference for flexibility as in Kreps (1979) and DLR (2001) into their temporal model. In addition, we allow for preference for timing to interact with preference for flexibility in a nontrivial way. For example, we allow for the possibility that preferences for timing are stronger when facing decision problems that offer flexibility in future choices. To illustrate the usefulness of these generalizations, we show in Section 1.3 that these new features have important implications for a broad class of mechanism design problems.

We describe the setting for our model in Section 2. The primitive of our model is a preference over lotteries over menus of lotteries. We interpret such preferences as arising from a two-period choice situation where in the first period, the individual chooses among lotteries over menus, and after a menu is realized, in the second period, she chooses a lottery out of the given menu. Our main results are utility representation and uniqueness theorems for the individual's behavior in the first period. We do not explicitly model the second-period choice out of the menus. A discussion of the second-period choices suggested by the utility representations is provided in Section 5.

We present our axioms and main results in Section 3. Let p denote a lottery over some set of alternatives Z, let A denote a menu of such lotteries, and let P denote a lottery over such menus. We show that any preference satisfying preference for early resolution of uncertainty (PERU) and our other axioms has the hidden action (HA) representation

$$V(P) = \int_{\mathcal{A}} \max_{\theta\in\Theta}\left(\int_{\Omega} \max_{p\in A} U(p,\omega;\theta)\pi(d\omega;\theta) - c(\theta)\right)P(dA). \tag{1}$$

We interpret (1) as follows. The individual is uncertain about her tastes over alternatives in Z. This uncertainty is modeled by a (subjective) state space Ω and a state-dependent expected-utility function U over $\triangle(Z)$. Before making a choice out of a menu A, i.e., after the resolution of the first-stage uncertainty but before the resolution of the second-stage uncertainty, the individual selects a hidden action $\theta \in \Theta$ that affects both her state-dependent utility function and the probability distribution over the states. After a state $\omega \in \Omega$ is realized, she learns her ex post utility function $U(\cdot,\omega;\theta)$, and chooses a lottery

$p \in A$ that maximizes it. A priori, faced with the menu A, the individual chooses her hidden action optimally by maximizing the ex ante expected value minus the cost $c(\theta)$ of the hidden action. She evaluates the lottery over menus P by taking the expectation over her payoff across different realizations of the menu A, giving (1). Intuitively, an HA representation satisfies PERU because the individual prefers to have objective uncertainty resolve in the first period so that she can choose her hidden action optimally. Our main representation theorem can be interpreted as saying that any preferences for early resolution of uncertainty can be represented as if the individual takes an unobserved (hidden) action between the resolution of the first- and second-period objective uncertainty.

It is well known that in models with state-dependent utility, the probability distribution over the states and the state-dependent utility function cannot be jointly identified from preferences.[1] For this reason, the parameters in an HA representation are not pinned down uniquely. To overcome this issue, we show in Lemma 1 that it is possible to transform the hidden actions in an HA representation into a reduced form by normalizing the state-dependent utility functions and using nonprobability measures. The use of nonprobability measures allows us to capture the combination of probability and magnitude of ex post utility jointly, resolving the nonuniqueness problem arising from state-dependent utility. We call this transformed representation the reduced-form hidden action (RFHA) representation. Theorem 1 shows the equivalence of the HA and the RFHA representations. We show in Theorem 3 that the parameters in our RFHA representation are uniquely identified from preference. We show in Lemma 2 that another benefit of thinking about hidden actions in reduced form is that it makes it possible to formalize a complementarity between menus and hidden actions in our representation.

In Section 4, we discuss certain special cases of our representation where the hidden actions take specific forms. In Section 4.1, we characterize the costly contemplation model of Ergin and Saver (2010a) as a special case where the hidden actions correspond to subjective signals/contemplation strategies over a subjective state space. We argue that in the costly contemplation model, a strict preference for early resolution only occurs in the presence of nondegenerate intermediate choice. In Section 4.2, we describe what is perhaps the simplest extension of the model of Kreps and Porteus (1978) that can accommodate a preference for flexibility. We call this special case the Kreps–Porteus–Dekel–Lipman–Rustichini (KPDLR) representation since it generalizes the main representations in Kreps and Porteus (1978) and DLR (2001). We show in Theorem 7 that a KPDLR representation inducing PERU is a special case of the HA representation: The hidden action acts as a common factor affecting the magnitude of the state-dependent utility while preserving the likelihood of the states and utility trade-offs across states. In a KPDLR representation, the preference for timing depends only on the utility values of the possible menus that could result from a two-stage lottery, not on the actual content of those menus. In particular, in contrast to the costly contemplation model, the presence or absence of intermediate choice has no direct effect on the preference for timing.

[1]See Kreps (1988) for a general discussion of the state-dependence issue. In Section 5.1, we provide an example of the nonuniqueness of the HA representation and a discussion of how incorporating second-period choice may aid in identification.

Hidden action representation	
$\int_{\mathcal{A}} \max_{\theta\in\Theta}(\int_{\Omega} \max_{p\in A} U(p,\omega;\theta)\pi(d\omega;\theta) - c(\theta))P(dA)$	
Special cases	Conditions
KPDLR with PERU	
– Probability measure independent of θ	$\pi(d\omega;\theta)=\pi(d\omega;\theta')\ \forall\theta,\theta'\in\Theta$
– Ranking of menus independent of θ	$U(p,\omega;\theta)=f(p,\omega)g(\theta)$
See Theorems 6.A and C, and 7	
DLR (2001)	
– No hidden action	Θ is a singleton
Machina (1984)	
– No state dependence	Ω is a singleton
Kreps and Porteus (1978, 1979) with PERU	
– No state dependence	Ω is a singleton
– Ranking of menus independent of θ	$U(p,\omega;\theta)=f(p)g(\theta)$
See Theorems 6.B, 6.C, and 7	
Ergin and Saver (2010a)	
– Probability measure independent of θ	$\pi(d\omega;\theta)=\pi(d\omega;\theta')\ \forall\theta,\theta'\in\Theta$
– Θ is a set of signals about the state	$\Theta=\mathbf{G}$ (set of sub-σ-algebras)
– U is the conditional expectation of $\bar{U}(\omega)\cdot p$	$U(p,\omega;\mathcal{G})=\mathbb{E}_{\pi}[\bar{U}\vert\mathcal{G}](\omega)\cdot p$
See Theorems 4 and 5	
Fixed (Stochastic) Ex Post Choice	
– Probability measure independent of θ	$\pi(d\omega;\theta)=\pi(d\omega;\theta')\ \forall\theta,\theta'\in\Theta$
– Ex post preferences over $\triangle(Z)$ independent of θ	$U(p,\omega;\theta)=f(p,\omega)g(\omega,\theta)$
See Section 5	

Table 1. Special cases of the hidden action representation.

Table 1 provides a summary of how the special cases discussed in Sections 1.1 and 4 relate to the HA representation in (1).

1.3 *A motivating example from mechanism design*

The choice objects we use in the paper naturally arise in mechanism design problems. In this section, we illustrate this connection and interpret our PERU axiom within the context of a mechanism design problem.

To make ideas concrete, focus on a simple school choice example with two schools, a and b, and two students, 1 and 2. Suppose student 1 has higher priority at school a and student 2 has higher priority at school b. The student-optimal matching mechanism gives the matches listed in Table 2 based on the reported preferences of the students.

Notice in particular that if student 2 reports a preference for school b, then given the priorities of the schools, student 1 is assigned to school a regardless of her preference. On the other hand, if student 2 reports a preference for school a, then student 1 is assigned to whichever school she ranks higher. Therefore, depending on the reported rankings of the other students and the priorities of the schools, there is a feasible set of

1's ranking	2's ranking	Matching
$a \succ_1 b$	$a \succ_2 b$	$(1, a)$, $(2, b)$
$a \succ_1 b$	$b \succ_2 a$	$(1, a)$, $(2, b)$
$b \succ_1 a$	$a \succ_2 b$	$(1, b)$, $(2, a)$
$b \succ_1 a$	$b \succ_2 a$	$(1, a)$, $(2, b)$

TABLE 2. Student-optimal matching mechanism.

2's ranking	1's feasible set
$a \succ_2 b$	$\{a, b\}$
$b \succ_2 a$	$\{a\}$

TABLE 3. Feasible choice sets for student 1.

schools for student 1 and she is assigned to her highest ranked school from this feasible set. The feasible sets for student 1 based on the reports of student 2 are summarized in Table 3.

This table concisely illustrates a key property of this mechanism: The ranking that student 1 submits has an impact on her outcome in some instances (when $a \succ_2 b$), but not in others (when $b \succ_2 a$). Suppose student 1 believes that with probability α, student 2 will submit the ranking $a \succ_2 b$ and with probability $1-\alpha$, student 2 will submit the ranking $b \succ_2 a$. Then, in the student-optimal stable matching mechanism, student 1 submits her ranking of a and b with the foresight that her choice of a versus b will be implemented with probability α, and with probability $1-\alpha$ she will be assigned a regardless of her reported ranking. Thus, submitting the ranking $a \succ_1 b$ results in a (for certain) and the ranking $b \succ_1 a$ results in the lottery $\alpha b + (1-\alpha)a$. This implies that her ranking of the alternatives (or, equivalently, her contingent plan from the sets $\{a, b\}$ and $\{a\}$) can be expressed within our framework as a choice from a set of distributions over outcomes $\{a, \alpha b + (1-\alpha)a\}$. In contrast, if student 1 learns the ranking of student 2 prior to submitting her own ranking, then with probability α, she chooses from the set $\{a, b\}$ and with probability $1-\alpha$, she chooses from the set $\{a\}$. When the decision problems for student 1 are formulated in this manner, our main axiom PERU corresponds precisely to student 1's desire to learn the ranking of student 2 prior to submitting her own ranking.

The student-optimal stable matching mechanism is dominant strategy incentive compatible for students. Therefore, in "standard" models, no student can benefit from learning the reports of the other students prior to submitting her own ranking. However, such preferences arise naturally in our model when students have access to unobservable actions that affect their payoffs.

There are many plausible examples of such hidden actions. Student 1 may prefer to learn her feasible choice set sooner to reduce her anxiety about the outcome or because there are other decisions in her life that she would like to condition on the outcome of the school match, such as a housing decision.[2] She could also find it difficult to assess

[2]The housing decision could be formalized in our HA model as a special case of the Machina (1984) representation in Table 1 where Ω is a singleton. Suppose that there are two hidden actions $\Theta = \{\theta_a, \theta_b\}$,

different strengths and weaknesses of the two schools.[3] In this case, knowing whether her feasible choice set is $\{a, b\}$ or $\{a\}$ prior to submitting her ranking is valuable to student 1, since she can put more effort into her decision when her submitted ranking is actually relevant and less effort when it is not. Under this interpretation, the preference for timing is motivated by the student wanting to avoid unnecessary contingent planning, i.e., investing the effort to rank schools that turn out not to be feasible. All of these examples of hidden actions are consistent with and can coexist in our model.

Different causes of preference for timing have very different implications, both for the overall structure of preferences and for the design of optimal mechanisms. For the preferences considered in Kreps and Porteus (1978), it is efficient to run the static mechanism at some optimal date (determined by the precise preferences for timing). For example, if students wish to learn the outcome of the school match prior to making housing decisions, running the same mechanism at an earlier date may lead to welfare gains. In contrast, if the preference for early resolution of uncertainty is purely to avoid unnecessary contingent planning, then running the same mechanism sooner would be of no use. In general, in our HA model, there is potential for efficiency gains associated with the use of dynamic mechanisms, since having agents act sequentially allows them to utilize information about the past actions of other agents when choosing their own action.

The applications of our model are not limited to the problem of school choice. The interpretation of the agents' reports as complete contingent plans naturally carries over to all economic environments in which a dominant strategy incentive-compatible direct revelation mechanism (like a Vickrey–Clarke–Groves (VCG) mechanism) is employed. Just as in the school choice example, for these mechanisms, every report of the other agents translates into a feasible set of outcomes that a particular agent can obtain from her different reports. If the mechanism is dominant strategy incentive-compatible, then it must always select the best feasible outcome for the agent according to her reported type. Therefore, submitting a report at the same time as other agents is equivalent to forming a contingent plan from the possible feasible sets, and agents with the preferences considered in this paper may again benefit from learning the reports of the other agents in advance of submitting their own.[4]

---

where θ_x corresponds to buying a house closer to school $x \in \{a, b\}$, and the cost function is identically zero. Since Ω is a singleton, we can suppress the state dependence of U and the probability measure π, and write the hidden action representation more compactly as $\int_{\mathcal{A}} \max_{x\in\{a,b\}}(\max_{p\in A} U(p, \theta_x))P(dA)$. When $U(b, \theta_b) > U(b, \theta_a) > U(a, \theta_a) > U(a, \theta_b)$, school b is preferred to school a independently of the housing choice. Since there is a benefit from conditioning the housing choice on the choice of the school, there is value to learning the feasible set prior to the choice of the hidden action.

[3]This can be formalized in our HA model as a special case of Ergin and Saver (2010a) in Table 1, where the probability measure π is independent of the hidden action. Suppose that there are two equally likely states $\Omega = \{\omega_a, \omega_b\}$ and two signals $\Theta = \{\theta_0, \theta_1\}$. Suppose that a is ahead in ω_a and b is ahead in ω_b, e.g., $\bar{U}(x, \omega_y) = 1$ if $x = y$ and $\bar{U}(x, \omega_y) = 0$ if $x \neq y$ for all $x, y \in \{a, b\}$. Suppose also that θ_0 is the uninformative signal generated by the trivial partition $\{\Omega\}$, θ_1 is the informative signal generated by the partition $\{\{\omega_a\}, \{\omega_b\}\}$, and $c(\theta_0) = 0 < c(\theta_1)$. The hidden action θ_0 gives $U(x, \omega_y, \theta_0) = \mathbb{E}_\pi[\bar{U}(x, \cdot)] = \frac{1}{2}$ and the hidden action θ_1 gives $U(x, \omega_y, \theta_1) = \bar{U}(x, \omega_y)$ for all $x, y \in \{a, b\}$. If $c(\theta_1) < \frac{1}{2}$, then the individual has a preference to learn the feasible set prior to selecting the signal.

[4]A simple example is the sealed-bid second-price auction with independent private values. Let b_i and b_{-i} denote the bid of agent i and the highest bid of the other agents, respectively. From the perspective of

In environments with monetary transfers, there are already results in the mechanism design literature illustrating the benefits associated with using dynamic mechanisms for several special cases of our general model. Athey and Segal (2013) provide an elegant construction of an efficient, budget-balanced, and Bayesian incentive-compatible dynamic mechanism in a setting where agents could have a very general set of hidden actions. When agents can engage in costly information acquisition about their private values (which corresponds to the special case of our model described in Section 4.1), Compte and Jehiel (2007) show that multistage auctions lead to higher revenues than sealed-bid auctions.[5]

2. Choice setting

Let Z be a finite set of alternatives, and let $\triangle(Z)$ denote the set of all probability distributions on Z, endowed with the Euclidean metric d and with generic elements denoted p, q, r. Let $\mathcal{A}$ denote the set of all nonempty and closed subsets of $\triangle(Z)$, endowed with the Hausdorff metric

$$d_h(A, B) = \max\Big\{\max_{p\in A}\min_{q\in B} d(p, q), \max_{q\in B}\min_{p\in A} d(p, q)\Big\}.$$

Elements of $\mathcal{A}$ are called menus, with generic menus denoted A, B, C. Let $\triangle(\mathcal{A})$ denote the set of all Borel probability measures on $\mathcal{A}$, endowed with the weak* topology and with generic elements denoted P, Q, R.[6] The primitive of the model is a binary relation $\succsim$ on $\triangle(\mathcal{A})$, representing the individual's preferences over lotteries over menus.

We interpret $\succsim$ as corresponding to the individual's choices in the first period of a two-period decision problem. In period 1, the individual first chooses a lottery P over menus. Then the uncertainty associated with this chosen lottery P resolves, returning a menu A. In the (unmodeled) period 2, the individual chooses a lottery p out of A and this lottery resolves, returning an alternative z. We will refer to the uncertainty associated with the resolution of P as the *first-stage uncertainty* and refer to the uncertainty associated with the resolution of p as the *second-stage uncertainty*. Although the period 2 choice is unmodeled, it will be important for the interpretation of the representations.[7]

agent i, each bid/report b_i corresponds to a complete contingent plan where she commits to buy the object at price b_{-i} if $b_i > b_{-i}$, and commits not to buy it if $b_i < b_{-i}$. A preference for timing driven by a desire to avoid contingent planning implies a preference by agent i to learn the price b_{-i} she faces before deciding whether or not to buy the object.

[5]These models involve agents who can take hidden actions in multiple stages, and the optimal dynamic mechanisms therefore also involve several stages. While our axiomatic analysis is restricted to two stages for tractability and expositional simplicity, the basic insights uncovered here are also useful for understanding multistage settings.

[6]Given a metric space X, the weak* topology on the set of all finite signed Borel measures on X is the topology where a net of signed measures $\{\mu_d\}_{d\in D}$ converges to a signed measure μ if and only if $\int_X f\mu_d(dx) \to \int_X f\mu(dx)$ for every bounded continuous function $f : X \to \mathbb{R}$.

[7]Since period 2 choice in our model is stochastic, incorporating it explicitly into the framework would involve a number of technical complications. We provide an informal discussion of how to approach the problem in Section 5, leaving the formal details for future research.

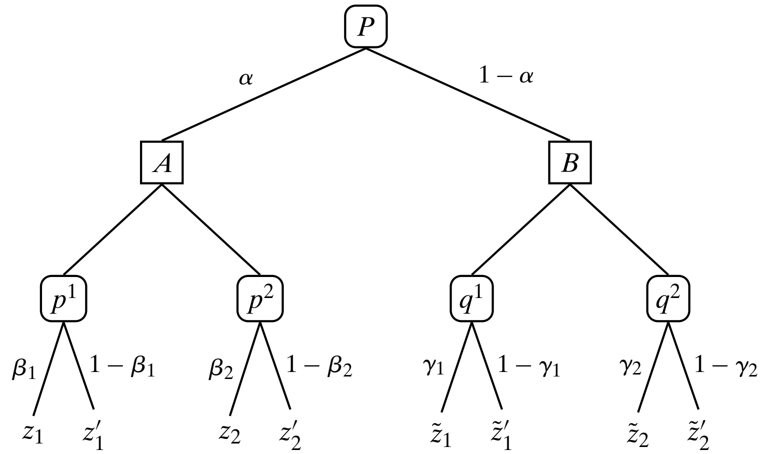

FIGURE 1. Decision tree for the lottery P.

For any $A \in \mathcal{A}$, let $\delta_A \in \triangle(\mathcal{A})$ denote the degenerate lottery that puts probability 1 on the menu A. Then $\alpha\delta_A + (1-\alpha)\delta_B$ denotes the lottery that puts probability α on the menu A and probability $1-\alpha$ on the menu B. Figure 1 illustrates such a lottery $P = \alpha\delta_A + (1-\alpha)\delta_B$ for the case of $A = \{p^1, p^2\}$ and $B = \{q^1, q^2\}$, where $p^i = \beta_i\delta_{z_i} + (1-\beta_i)\delta_{z_i'}$ and $q^i = \gamma_i\delta_{\tilde{z}_i} + (1-\gamma_i)\delta_{\tilde{z}_i'}$. In this figure, nodes with rounded edges are those at which nature acts, and square nodes are those at which the individual makes a decision.

Our framework is a special case of that of Kreps and Porteus (1978), with only two periods and no consumption in period 1.[8] As in Kreps and Porteus (1978), we refer to a lottery $P \in \triangle(\mathcal{A})$ over menus as a *temporal lottery* if P returns a singleton menu with probability 1. An individual facing a temporal lottery makes no choice in period 2 between the resolution of first and second stages of the uncertainty. Note that the set of temporal lotteries can be naturally associated with $\triangle(\triangle(Z))$.

For any $A, B \in \mathcal{A}$ and $\alpha \in [0,1]$, the convex combination of these two menus is defined by $\alpha A + (1-\alpha)B \equiv \{\alpha p + (1-\alpha)q : p \in A \text{ and } q \in B\}$. Let $\mathrm{co}(A)$ denote the convex hull of the menu A. Finally, for any continuous function $V : \mathcal{A} \to \mathbb{R}$ and $P \in \triangle(\mathcal{A})$, we let $\mathbb{E}_P[V]$ denote the expected value of V under the lottery P, i.e., $\mathbb{E}_P[V] = \int_{\mathcal{A}} V(A)P(dA)$.

3. THE GENERAL REPRESENTATION

3.1 *Axioms*

We will impose the following set of axioms in all the representation results in the paper. Therefore, it will be convenient to refer to them altogether as Axiom 1.

AXIOM 1. *(i)* Weak order*: The binary relation $\succsim$ is complete and transitive.*

(ii) Continuity*: The upper and lower contour sets, $\{P \in \triangle(\mathcal{A}) : P \succsim Q\}$ and $\{P \in \triangle(\mathcal{A}) : P \precsim Q\}$, are closed in the weak* topology.*

[8]This framework was also used in Epstein and Seo (2009) and in Section 4 of Epstein et al. (2007).

(iii) First-stage independence*: For any $P, Q, R \in \triangle(\mathcal{A})$ and $\alpha \in (0,1)$,*

$$P \succ Q \quad \Rightarrow \quad \alpha P + (1-\alpha)R \succ \alpha Q + (1-\alpha)R.$$

(iv) L-continuity*: There exist $A^*, A_* \in \mathcal{A}$ and $M \geq 0$ such that for every $A, B \in \mathcal{A}$ and $\alpha \in [0,1]$ with $\alpha \geq M d_h(A,B)$,*

$$(1-\alpha)\delta_A + \alpha\delta_{A^*} \succsim (1-\alpha)\delta_B + \alpha\delta_{A_*}.$$

Axioms 1(i) and (ii) are standard. Axiom 1(iii) is the von Neumann–Morgenstern independence axiom imposed with respect to the first-stage uncertainty. Axioms 1(i)–(iii) ensure that there exists a continuous function $V : \mathcal{A} \to \mathbb{R}$ such that $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$. Given Axioms 1(i)–(iii), Axiom 1(iv) is a technical condition implying the Lipschitz continuity of V.[9]

Kreps and Porteus (1978) defined preference for early resolution of uncertainty using temporal lotteries. Formally, their preference for early resolution of uncertainty (PERU) axiom states that for any $p, q \in \triangle(Z)$ and $\alpha \in [0,1]$,

$$\alpha\delta_{\{p\}} + (1-\alpha)\delta_{\{q\}} \succsim \delta_{\{\alpha p + (1-\alpha)q\}}. \tag{2}$$

In the temporal lottery $\alpha\delta_{\{p\}} + (1-\alpha)\delta_{\{q\}}$, uncertainty regarding whether lottery p or q is selected resolves in period 1. In the temporal lottery $\delta_{\{\alpha p+(1-\alpha)q\}}$, the same uncertainty resolves in period 2.[10] PERU requires a weak preference for the first temporal lottery. One can similarly define preference for late resolution of uncertainty (PLRU) and indifference to the timing of resolution of uncertainty (ITRU) axioms. However, since our main focus is on PERU, we relegate discussion of the representations for PLRU to Appendix B.

Figure 2 illustrates such temporal lotteries in the special case where $p = \delta_z$ and $q = \delta_{\tilde{z}}$ for some $z, \tilde{z} \in Z$. In this figure, nodes with rounded edges are those at which nature acts, and rectangular nodes are those at which the individual makes a decision. Since the trees in this figure correspond to temporal lotteries, the action nodes for the individual are always degenerate. The temporal lottery $\alpha\delta_{\{\delta_z\}} + (1-\alpha)\delta_{\{\delta_{\tilde{z}}\}}$ corresponds to the first tree in Figure 2, in which the uncertainty about whether alternative z or $\tilde{z}$ will be selected resolves in period 1. The temporal lottery $\delta_{\alpha\{\delta_z\}+(1-\alpha)\{\delta_{\tilde{z}}\}}$ corresponds to the second tree in Figure 2, in which the uncertainty about whether z or $\tilde{z}$ will be selected resolves in period 2.

Kreps and Porteus (1978) impose other axioms that tie the preference for early resolution of uncertainty for general two-stage decision problems to the preference for early resolution of uncertainty on temporal lotteries. Since we make weaker overall assumptions on preferences, we adapt their axiom to be explicit about the preferences being imposed on lotteries involving nondegenerate choices.[11]

[9] In models with preferences over menus of lotteries, related L-continuity axioms were used by Dekel, Lipman, Rustichini, and Sarver (2007, henceforth DLRS), Sarver (2008), and Ergin and Saver (2010a).

[10] In both temporal lotteries, the remaining uncertainty, i.e., the outcome of p conditional on p being selected and the outcome of q conditional on q being selected, is also resolved in period 2.

[11] Note that while our preference for early resolution of uncertainty axiom is stronger than that explicitly stated by Kreps and Porteus (1978), Axiom 2 is implied by its temporal lottery counterpart when the

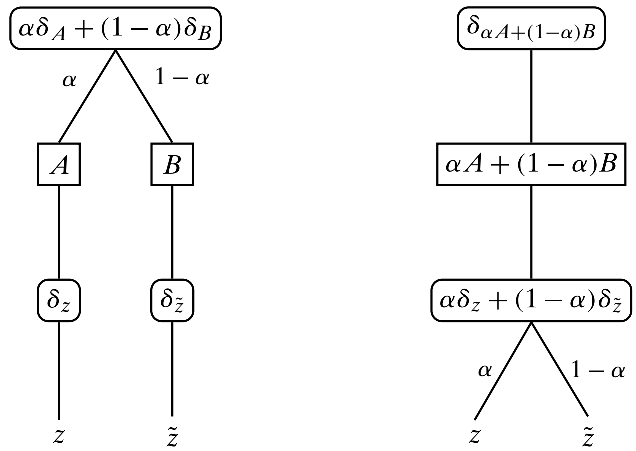

FIGURE 2. Illustration of the timing of resolution of uncertainty for temporal lotteries: $A = \{\delta_z\}$ and $B = \{\delta_{\tilde{z}}\}$.

AXIOM 2 (Preference for early resolution of uncertainty (PERU)). *For any $A, B \in \mathcal{A}$ and $\alpha \in (0,1)$,*

$$\alpha\delta_A + (1-\alpha)\delta_B \succsim \delta_{\alpha A+(1-\alpha)B}.$$

In the early resolution lottery $\alpha\delta_A + (1-\alpha)\delta_B$, any uncertainty regarding the feasible set resolves in period 1, in particular, before the individual makes a choice from the realized menu. In the late resolution lottery $\delta_{\alpha A+(1-\alpha)B}$, the individual learns nothing in period 1 and then makes a choice from the menu $\alpha A + (1-\alpha)B$. We interpret this menu as the set of all contingent plans from the menus A and B (or, more precisely, the distributions over outcomes resulting from those contingent plans). To understand this interpretation, suppose the individual is asked to make a contingent plan $(p, q) \in A \times B$, where p will be implemented if the realized menu is A and q will be implemented in the case of B. Since A will be the relevant menu with probability α, this contingent plan induces the distribution over outcomes $\alpha p + (1-\alpha)q \in \alpha A + (1-\alpha)B$.

With this interpretation in mind, late resolution of uncertainty corresponds to learning nothing in the period 1 and then making a contingent plan (from the yet unrealized choice sets) that will be carried out after this uncertainty resolves in period 2. Therefore, timing of resolution of uncertainty can be broken into two components in our model:

1. *Absolute timing*: Whether the individual gets information sooner or later.
2. *Relative timing*: Whether the individual gets information prior to committing to a plan of action or not.

other axioms of Kreps and Porteus (1978) are imposed. It is also worth noting that other authors have used stronger versions of the preference for early resolution of uncertainty axiom to relax other assumptions on the preferences. For example, to study recursive nonexpected-utility models over temporal lotteries, Grant et al. (1998, 2000) introduced a stronger version of (2) that, roughly speaking, requires individuals to prefer when the resolution of the first-stage uncertainty is more informative in the sense of Blackwell.

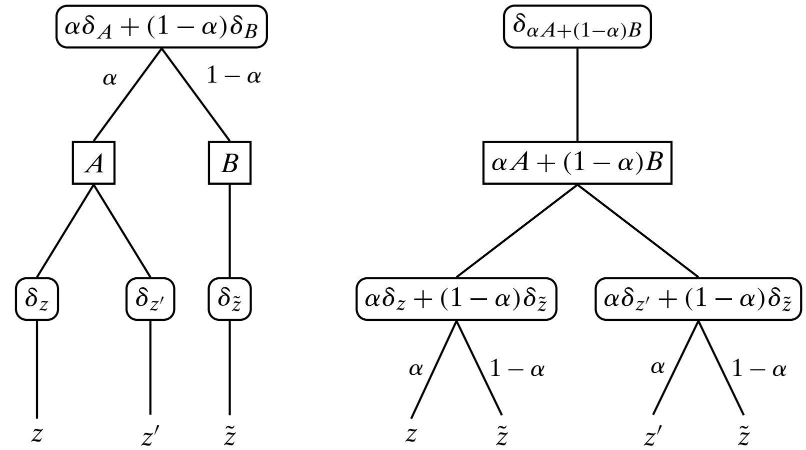

FIGURE 3. Illustration of the timing of resolution of uncertainty for nontemporal lotteries: $A = \{\delta_z, \delta_{z'}\}$ and $B = \{\delta_{\tilde{z}}\}$.

If the presence or absence of intermediate choice is inconsequential for the preference for timing (as in the model of Kreps and Porteus 1978), we can infer that only absolute timing is important to the individual. Alternatively, if the preference for timing changes in the presence of intermediate choice, then relative timing is also relevant. By taking into account both absolute and relative timing of uncertainty, we can model novel issues such as difficulty in making complex contingent decisions.

When considering the timing of uncertainty relative to choice, it is important to keep in mind the potential instrumental value of information. As illustrated by the simple consumption/savings example in Section 1.1, changing the timing of information relative to choice has the potential to alter the individual's ability to condition her actions on the realization of uncertainty. However, this well understood interaction between information and choice is not at work in our preference for early resolution of uncertainty axiom. The use of contingent plans in our comparison of late versus early resolution of uncertainty ensures that the individual's ability to condition her choices on the realized set is unaffected by the timing of resolution of uncertainty. Since the distributions over final outcomes available to the individual are the same in the case of early or late resolution, the only difference is whether she must commit to a plan of action prior to the resolution of uncertainty.

Figure 3 illustrates timing of resolution of uncertainty in the case where $A = \{\delta_z, \delta_{z'}\}$ and $B = \{\delta_{\tilde{z}}\}$. The lottery $\alpha\delta_A + (1-\alpha)\delta_B$ corresponds to the first tree in Figure 3, in which the uncertainty about whether the choice set will be A or B resolves in period 1, before the individual makes her choice from the realized menu. The lottery $\delta_{\alpha A+(1-\alpha)B}$ corresponds to the second tree in Figure 3, in which the individual's period 2 choice is made prior to the resolution of uncertainty regarding whether her choice from A or B will be implemented. In this tree, the lottery $\alpha\delta_z + (1-\alpha)\delta_{\tilde{z}}$ can be interpreted as a contingent plan where the individual commits to choosing δ_z if A is the realized choice set and $\delta_{\tilde{z}}$ if B is the realized choice set. Similarly, $\alpha\delta_{z'} + (1-\alpha)\delta_{\tilde{z}}$ corresponds to making a contingent choice of $\delta_{z'}$ from the menu A.

The final axiom for our general model is a standard monotonicity axiom, which requires a weak preference for larger menus.

Axiom 3 (Monotonicity). *For any $A, B \in \mathcal{A}$, $A \subset B$ implies $\delta_B \succsim \delta_A$.*

Kreps (1979) and DLR (2001) used this axiom to capture a preference for flexibility. For example, if the individual is uncertain of whether she will prefer to choose lottery p or q in period 2, then in period 1 she may strictly prefer to retain the flexibility of $\delta_{\{p,q\}}$ rather than committing to either $\delta_{\{p\}}$ or $\delta_{\{q\}}$.

In contrast, if the individual anticipates that her future choices will be inconsistent with her current preferences, she may strictly prefer to commit to a smaller menu. For example, Gul and Pesendorfer (2001) and DLR (2009) relaxed monotonicity in a menu-choice setting so as to model temptation and costly self-control. Our focus is, instead, on the interaction between preferences for flexibility and timing, so we impose monotonicity throughout the main text. However, in Appendix B, we describe a generalization of our main representation to nonmonotone preferences, which can be used as a starting point for future research on incorporating temptation into our temporal model.

It has also been suggested that preferences for early or late resolution of uncertainty could also arise due to anticipatory feelings or anxiety. Our hidden action representation is, in principle, consistent with such an interpretation; for example, anticipating a particular level of consumption could be thought of as a hidden action on the part of the individual (see Sarver 2014). However, our axioms are inconsistent with several earlier models of anticipatory feelings in the literature (e.g., Caplin and Leahy 2001 and Epstein 2008) because of our assumption of monotonicity. Loosely speaking, Caplin and Leahy (2001) and Epstein (2008) assume that anticipation/anxiety has a greater impact on utility in early stages than in later, which causes the individual's ranking of lotteries to change over time. If the individual correctly foresees that she will be dynamically inconsistent in this way, then she will strictly prefer to commit herself to a particular lottery at the first stage and, hence, will violate monotonicity.

3.2 *The hidden action representation*

Note that expected-utility functions on $\triangle(Z)$ are equivalent to vectors in $\mathbb{R}^Z$ by associating each expected-utility function with its values for sure outcomes. We therefore use the notation $u(p)$ and $u \cdot p$ interchangeably for any $u \in \mathbb{R}^Z$. Next, we give the formal definition of the hidden action (HA) representation discussed in the Introduction.

Definition 1. A *hidden action (HA) representation* is a tuple $((\Omega, \mathcal{F}, \pi), \Theta, U, c)$, where $(\Omega, \mathcal{F})$ is a measurable space, Θ is a set, $\pi(\cdot;\theta)$ is a probability measure on $(\Omega, \mathcal{F})$ for every $\theta \in \Theta$, $U : \Omega \times \Theta \to \mathbb{R}^Z$ is a bounded function such that $U(\cdot, \theta)$ is an $\mathcal{F}$-measurable random vector for every $\theta \in \Theta$, $U(p, \omega; \theta) := U(\omega, \theta) \cdot p$, and $c : \Theta \to \mathbb{R}$ is a function such that $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$, where $V : \mathcal{A} \to \mathbb{R}$ is defined by

$$V(A) = \max_{\theta \in \Theta}\left(\int_\Omega \max_{p \in A} U(p, \omega; \theta)\pi(d\omega; \theta) - c(\theta)\right) \tag{3}$$

and the maximization in (3) has a solution for every $A \in \mathcal{A}$.[12]

The interpretation of the HA representation is as follows. The individual is uncertain about her tastes over $\triangle(Z)$. This uncertainty is modeled by the state space Ω and a state-dependent utility function U. Before making a choice out of a menu A, i.e., after the resolution of the first-stage uncertainty but before the resolution of the second-stage uncertainty, the individual is able to select a hidden action $\theta \in \Theta$ that affects both her state-dependent utility function and the probability distribution over the states. After she chooses $\theta \in \Theta$ and a state $\omega \in \Omega$ is realized, she learns her ex post utility function $U(\cdot, \omega; \theta)$, and chooses a lottery $p \in A$ that maximizes it. A priori, faced with the menu A, the individual chooses her hidden action optimally by maximizing the ex ante expected value minus the cost $c(\theta)$ of the hidden action, giving (3).

We postpone more concrete interpretations of the set of hidden actions and costs to the discussion of the special cases in the following section. Table 1 in the Introduction provides a summary of how these special cases relate to the HA representation.

The HA representation is not unique in the sense that two different sets of HA parameters can lead to the same value function V in (3). The lack of identification of probabilities is a common issue in models with state-dependent utility.[13] We next argue that it is possible to transform the HA representation into a reduced form where the parameters will be identified from preference. Our approach generalizes that in Ergin and Saver (2010a).

We define the set of *normalized (nonconstant) expected-utility functions* on $\triangle(Z)$ to be

$$\mathcal{U} = \left\{ u \in \mathbb{R}^Z : \sum_{z \in Z} u_z = 0, \sum_{z \in Z} u_z^2 = 1 \right\}.^{14}$$

Modulo an affine transformation, $\mathcal{U}$ contains all possible ex post expected-utility functions. The following lemma illustrates how to reexpress the expectation term in (3) uniquely by normalizing the state-dependent utility functions and using nonprobability measures μ that capture the combination of the probability and magnitude (cardinality) of ex post utility.

[12] An argument similar to the one in Appendix A in Ergin and Saver (2010a) can be used to show that the boundedness of U and the $\mathcal{F}$-measurability of $U(\cdot, \theta)$ imply that the integral $\int_\Omega \max_{p \in A} U(p, \omega; \theta)\pi(d\omega; \theta)$ is well defined and finite for every $A \in \mathcal{A}$ and $\theta \in \mathbf{\Theta}$. Also, for simplicity, we directly assume that the outer maximization in (3) has a solution. An alternative approach would be to impose topological assumptions on the parameters that would guarantee the existence of a maximum, for instance, assuming that Θ is a compact topological space, Ω is a metric space, $\mathcal{F}$ is the Borel σ-algebra on Ω, $\pi : \Theta \to \triangle(\Omega)$ is continuous, where $\triangle(\Omega)$ denotes the set of probability measures on $(\Omega, \mathcal{F})$ endowed with the weak* topology, $U(\cdot, \theta) : \Omega \to \mathbb{R}^Z$ is continuous for every $\theta \in \Theta$, the collection of functions $\{U(\omega, \cdot)\}_{\omega \in \Omega}$ is equicontinuous in θ, and c is lower semicontinuous. Yet another alternative approach would be to assume that c is bounded below and replace the outer maximization in (3) with a supremum.

[13] See Kreps (1988) for a general discussion of the state-dependence issue, and Section 3 of Ergin and Saver (2010a) and Section 5.1 of this paper for discussions specific to this setting.

[14] We endow the set of all finite Borel measures on $\mathcal{U}$ with the weak* topology (see footnote 6).

LEMMA 1. *For any* $((\Omega, \mathcal{F}, \pi), \Theta, U)$ *as in an HA representation and* $\theta \in \Theta$*, there exists a unique finite Borel measure* μ_θ *on* $\mathcal{U}$ *and scalar* β_θ *such that for all* $A \in \mathcal{A}$*,*[15]

$$\int_\Omega \max_{p \in A} U(p, \omega; \theta)\pi(d\omega; \theta) = \int_\mathcal{U} \max_{p \in A} u(p)\mu_\theta(du) + \beta_\theta.$$

Conversely, for any compact set $\mathcal{M}$ *of finite Borel measures on* $\mathcal{U}$*, there is* $((\Omega, \mathcal{F}, \pi), \Theta, U)$ *as in an HA representation where* $\Theta = \mathcal{M}$*, such that for all* $\mu \in \mathcal{M}$ *and* $A \in \mathcal{A}$*,*

$$\int_\mathcal{U} \max_{p \in A} u(p)\mu(du) = \int_\Omega \max_{p \in A} U(p, \omega; \mu)\pi(d\omega; \mu).$$

PROOF. Since this lemma follows from the same arguments used to prove Lemma 1 in Ergin and Saver (2010a), we only provide the key steps. To prove the first claim, note that for every $\omega \in \Omega$, there exist $a_\omega \geq 0$, $b_\omega \in \mathbb{R}$, and $u_\omega \in \mathcal{U}$ such that $U(\omega, \theta) = a_\omega u_\omega + b_\omega$. Let $\beta_\theta = \int_\Omega b_\omega \pi(d\omega; \theta)$, and define a Borel measure μ_θ by $\mu_\theta(E) = \int_{\{\omega \in \Omega : u_\omega \in E\}} a_\omega \pi(d\omega; \theta)$ for a measurable set $E \subset \mathcal{U}$. Using a standard change of variables, it follows that for every $A \in \mathcal{A}$,

$$\begin{aligned}\int_\Omega \max_{p \in A} U(p, \omega; \theta)\pi(d\omega; \theta) &= \int_\Omega a_\omega \max_{p \in A} u_\omega(p)\pi(d\omega; \theta) + \int_\Omega b_\omega \pi(d\omega; \theta) \\ &= \int_\mathcal{U} \max_{p \in A} u(p)\mu_\theta(du) + \beta_\theta.\end{aligned}$$

Intuitively, the magnitude of each utility function $U(\omega, \theta)$ is incorporated into the measure of the corresponding u_ω.

To prove the converse, let $\Omega = \mathcal{U}$, $\mathcal{F}$ be the Borel σ-algebra on Ω and let $\Theta = \mathcal{M}$. Let $\mu \in \mathcal{M}$ and $\lambda \equiv \mu(\mathcal{U}) \geq 0$. If $\lambda = 0$, define $U(\omega, \mu) = 0$ for all $\omega \in \Omega$ and let $\pi(\cdot, \mu)$ be an arbitrary probability measure on $(\Omega, \mathcal{F})$. Otherwise, define $U(\omega, \mu) = \lambda\omega$ for all $\omega \in \Omega$ and $\pi(E, \mu) = \mu(E)/\lambda$ for any measurable set $E \subset \Omega$. Heuristically, given the hidden action μ, the probability measure $\pi(\cdot, \mu)$ puts weight $\mu(\omega)/\lambda$ on the state where the ex post utility function is $\lambda\omega$. Therefore,

$$\int_\mathcal{U} \max_{p \in A} u(p)\mu(du) = \frac{1}{\lambda}\int_\mathcal{U} \max_{p \in A} \lambda u(p)\mu(du) = \int_\Omega \max_{p \in A} U(p, \omega; \mu)\pi(d\omega; \mu). \qquad \square$$

The first part of Lemma 1 shows that each hidden action θ in an HA representation corresponds to a unique measure μ_θ over $\mathcal{U}$. Note that the normalization of the ex post utility functions in $\mathcal{U}$ is necessary for obtaining the uniqueness of μ_θ. The converse shows that any collection $\mathcal{M}$ of measures over $\mathcal{U}$ might be given a (nonunique) probabilistic interpretation through an HA representation.

There is a certain type of complementarity between menus and hidden actions in the HA representation. Intuitively, "increasing" θ will have a bigger impact on larger menus. Besides the identification of the representation, another benefit of thinking about hidden actions in reduced form is that we can state this complementarity formally using the natural partial order on reduced-form hidden actions.

[15]Note that the constant β_θ can be absorbed into the function c when transforming the HA representation into reduced form.

Lemma 2. *Let $((\Omega, \mathcal{F}, \pi), \Theta, U, c)$ be an HA representation. For every menu $A \in \mathcal{A}$ and hidden action $\theta \in \Theta$, define*

$$V(A|\theta) = \int_{\Omega} \max_{p \in A} U(p, \omega; \theta)\pi(d\omega; \theta).$$

For each $\theta \in \Theta$, let μ_θ be the unique finite Borel measure on $\mathcal{U}$ from Lemma 1, and consider the binary relation $\trianglelefteq$ on Θ defined by

$$\theta \trianglelefteq \theta' \quad \Leftrightarrow \quad \mu_\theta \leq \mu_{\theta'}.$$

Then $V(\cdot|\cdot)$ has increasing differences in A and θ, i.e., if $A \subset B$ and $\theta \trianglelefteq \theta'$, then

$$V(A|\theta') - V(A|\theta) \leq V(B|\theta') - V(B|\theta).$$

Proof. By Lemma 1, the right-hand side minus the left-hand side of the desired inequality can be expressed as

$$V(B|\theta') - V(B|\theta) - [V(A|\theta') - V(A|\theta)] = \int_{\mathcal{U}} \Big[\max_{p \in B} u(p) - \max_{p \in A} u(p)\Big][\mu_{\theta'} - \mu_\theta](du) \geq 0,$$

where the integrand in the brackets in nonnegative because $A \subset B$, and the measure $[\mu_{\theta'} - \mu_\theta]$ is nonnegative because $\theta \trianglelefteq \theta'$. $\square$

Motivated by Lemmas 1 and 2, we now introduce our reduced-form representation:

Definition 2. A *reduced-form hidden action* (RFHA) representation is a pair $(\mathcal{M}, c)$ consisting of a compact set of finite Borel measures $\mathcal{M}$ on $\mathcal{U}$ and a lower semicontinuous function $c : \mathcal{M} \to \mathbb{R}$ such that the following statements hold:

(i) We have $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$, where $V : \mathcal{A} \to \mathbb{R}$ is defined by

$$V(A) = \max_{\mu \in \mathcal{M}} \left(\int_{\mathcal{U}} \max_{p \in A} u(p)\mu(du) - c(\mu) \right). \tag{4}$$

(ii) The set $\mathcal{M}$ is *minimal*: For any compact proper subset $\mathcal{M}'$ of $\mathcal{M}$, the function V' obtained by replacing $\mathcal{M}$ with $\mathcal{M}'$ in (4) is different from V.

The reduced-form formulation in Definition 2 will have the important benefit of allowing for the unique identification of the set of reduced-form available actions $\mathcal{M}$ and their costs c from preference. It will also simplify the mathematical statement of some results. Note that in (4), it is possible to enlarge the set of actions by adding a new action μ to the set $\mathcal{M}$ at a prohibitively high cost $c(\mu)$ without affecting the equation. Therefore, so as to identify $(\mathcal{M}, c)$ from the preference, we also impose an appropriate minimality condition on the set $\mathcal{M}$.

The following result formalizes the interpretation of the RFHA representation as a reduced form of the HA representation.

THEOREM 1. *Let $V : \mathcal{A} \to \mathbb{R}$. Then there exists an HA representation such that V is given by (3) if and only if there exists an RFHA representation such that V is given by (4).*

We are now ready to state our main representation result.

THEOREM 2. *The preference $\succsim$ has an RFHA representation if and only if it satisfies Axiom 1, PERU, and monotonicity.*

The proof of Theorem 2 is contained in Appendix B. In that section, we also present several extensions that are outside the main focus of our analysis but may be useful for subsequent research: First, we show it is possible to relax the assumption of monotonicity if signed measures are permitted in the RFHA representation. Second, we describe a dual version of the RFHA representation that corresponds to preference for late resolution of uncertainty (PLRU). The dual representation is related to a representation considered by Epstein et al. (2007) in the setting of menus of lotteries. We discuss the connection between the two representations in Appendix B.

Using Theorems 1 and 2, we obtain the following HA representation result as a corollary.

COROLLARY 1. *The preference $\succsim$ has an HA representation if and only if it satisfies Axiom 1, PERU, and monotonicity.*

The special cases of an RFHA representation satisfying indifference to timing of resolution of uncertainty ($\alpha\delta_A + (1-\alpha)\delta_B \sim \delta_{\alpha A+(1-\alpha)B}$) are those where $\mathcal{M}$ is a singleton. For those cases, the constant cost can be dropped from (4), leading to an analogue of DLR (2001)'s additive representation in which the individual reduces compound lotteries.

We next give a brief intuition about Theorem 2. Our axioms guarantee the existence of a Lipschitz continuous function $V : \mathcal{A} \to \mathbb{R}$ such that $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$. In terms of this expected utility representation, it is easy to see that PERU corresponds to convexity of V. Now consider the following axiom.

AXIOM 4 (Indifference to randomization (IR)). *For every $A \in \mathcal{A}$, $\delta_A \sim \delta_{\mathrm{co}(A)}$.*

Axiom 4 was introduced in DLR (2001). It is justified if the individual choosing from the menu A in period 2 can also randomly select an alternative from the menu, for example, by flipping a coin. In that case, the menus A and $\mathrm{co}(A)$ offer the same set of options, and, hence, they are identical from the perspective of the individual. Lemma 5 in Appendix B shows that weak order, continuity, PERU, and monotonicity imply IR.

Given IR, we can restrict attention to convex menus. Moreover, the set $\mathcal{A}^c$ of convex menus can be mapped one-to-one to a set of continuous functions Σ known as the support functions, preserving the metric and the linear operations. Therefore, by using the property $V(\mathrm{co}(A)) = V(A)$ implied by IR and mimicking the construction in DLR (2001), V can be thought of as a function defined on the subset Σ of the Banach space $C(\mathcal{U})$ of continuous real-valued functions on $\mathcal{U}$. This allows us to apply a variation of the classic duality principle that convex functions can be written as the supremum of affine functions lying below them.[16] Then we apply the Riesz representation theorem to write each such continuous affine function as an integral against a measure μ minus a scalar $c(\mu)$. Finally, imposing monotonicity guarantees that all measures in the RFHA representation are positive.

We show that the uniqueness of the RFHA representation follows from the affine uniqueness of V and a result about the uniqueness of the dual representation of a convex function from the theory of conjugate convex functions (see Theorem 10 in Appendix A). A similar application of the duality and uniqueness results can be found in Ergin and Saver (2010a). In terms of the HA representation, Theorem 3 can be interpreted as an identification of the equivalence classes of HA representations that lead to the same choice behavior.

THEOREM 3. *If $(\mathcal{M}, c)$ and $(\mathcal{M}', c')$ are two RFHA representations for $\succsim$, then there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $\mathcal{M}' = \alpha\mathcal{M}$ and $c'(\alpha\mu) = \alpha c(\mu) + \beta$ for all $\mu \in \mathcal{M}$.*

4. Special cases

4.1 *Costly contemplation*

Recall that a choice out of the convex combination menu $\alpha A + (1 - \alpha)B$ can be interpreted as a complete contingent plan out of the two menus A and B: Each lottery $\alpha p + (1 - \alpha)q \in \alpha A + (1 - \alpha)B$ is identical to a pair of choices $p \in A$ and $q \in B$, where after the individual chooses (p, q), p is selected with probability α and q is selected with probability $1 - \alpha$. Therefore, PERU can be naturally attributed to a desire to avoid making complete contingent plans. Note, however, that a pure desire to avoid contingent planning is a special kind of PERU. For instance, when the menus A and B are singletons so that the contingent planning problem faced in $\alpha A + (1 - \alpha)B$ is trivial, there is no reason for an individual who is averse to contingent planning to prefer $\alpha\delta_A + (1 - \alpha)\delta_B$ over $\delta_{\alpha A + (1-\alpha)B}$. In particular, if the driving force underlying an individual's PERU is solely an aversion to contingent planning, then it is natural to observe indifference to timing of resolution of uncertainty over temporal lotteries.

In Ergin and Saver (2010a), we studied preferences exhibiting aversion to contingent planning in the simpler framework of preferences over menus of lotteries. We obtained a representation for such preferences that can be interpreted in terms of costly contemplation. The following is the natural extension of that representation to the current framework of lotteries over menus.

DEFINITION 3. A *costly contemplation* (CC) representation is a tuple $((\Omega, \mathcal{F}, \pi), \mathbf{G}, U, c)$, where $(\Omega, \mathcal{F}, \pi)$ is a probability space, $\mathbf{G}$ is a collection of sub-σ-algebras of $\mathcal{F}$, $U : \Omega \to \mathbb{R}^Z$ is a Z-dimensional, $\mathcal{F}$-measurable, and integrable random vector, and $c : \mathbf{G} \to \mathbb{R}$ is a function, such that $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$, where $V : \mathcal{A} \to \mathbb{R}$ is defined by

$$V(A) = \max_{\mathcal{G} \in \mathbf{G}} \Big(\mathbb{E}_\pi \Big[\max_{p \in A} \mathbb{E}_\pi [U | \mathcal{G}] \cdot p \Big] - c(\mathcal{G}) \Big), \tag{5}$$

[16]See Rockafellar (1970), Phelps (1993), and Appendix A of the current paper for variations of this duality result.

and the maximization in (5) has a solution for every $A \in \mathcal{A}$.[17]

The interpretation of the CC representation is as follows. The individual is uncertain about her tastes over $\triangle(Z)$. This uncertainty is modeled by a probability space $(\Omega, \mathcal{F}, \pi)$ and a state-dependent expected-utility function U over $\triangle(Z)$. Before making a choice out of a menu A, the individual is able to engage in contemplation so as to resolve some of this uncertainty. Contemplation strategies are modeled as signals about the state or, more compactly, as a collection $\mathbf{G}$ of σ-algebras generated by these signals. If the individual carries out the contemplation strategy $\mathcal{G}$, she is able to update her expected-utility function using her information $\mathcal{G}$ and choose a lottery p in A maximizing her conditional expected-utility $\mathbb{E}_\pi[U|\mathcal{G}] \cdot p$. Faced with the menu A, the individual chooses her contemplation strategy optimally by maximizing the ex ante value minus the cost $c(\mathcal{G})$ of contemplation, giving (5). Note that the CC formula is mathematically identical to a standard costly information acquisition problem. The difference is that the parameters $((\Omega, \mathcal{F}, \pi), \mathbf{G}, U, c)$ of the CC representation are subjective in the sense that they are not directly observable, but instead must be elicited from the individual's preferences.[18]

Theorem 2 from Ergin and Saver (2010a) can be applied to the current setting to show that a CC representation can be written in reduced form as an HA representation satisfying a consistency condition.[19]

Theorem 4 (Ergin and Saver 2010a). *Let $V : \mathcal{A} \to \mathbb{R}$. Then the following statements are equivalent:*

(i) There exists a CC representation such that V is given by (5).

(ii) There exists an RFHA representation $(\mathcal{M}, c)$ such that V is given by (4) and $\mathcal{M}$ satisfies consistency*:*

$$\forall \mu, \nu \in \mathcal{M} \textit{ and } \forall p \in \triangle(Z): \quad \int_{\mathcal{U}} u(p)\mu(du) = \int_{\mathcal{U}} u(p)\nu(du).$$

Therefore, consistency is key for the interpretation of the RFHA representation as a subjective information acquisition problem. The intuition for how a CC representation

[17]We showed in Ergin and Saver (2010a) that the integrability of U implies that the term $\mathbb{E}_\pi[\max_{p\in A} \mathbb{E}_\pi[U|\mathcal{G}] \cdot p]$ is well defined and finite for every $A \in \mathcal{A}$ and $\mathcal{G} \in \mathbf{G}$. For simplicity, we directly assume that the outer maximization in (5) has a solution instead of making topological assumptions on $\mathbf{G}$ to guarantee the existence of a maximum. An alternative approach that does not require this indirect assumption on the parameters of the representation would be to assume that c is bounded below, and to replace the outer maximization in (5) with a supremum, in which case all of our results would carry over.

[18]The costly contemplation representation in (5) is similar to the functional form considered in Ergin (2003), where the primitive is a preference over menus taken from a finite set of alternatives. Ortoleva (2013) also considered a related model of costly thinking using slightly different primitives. The main conceptual distinction from our model is that he considered an individual who may choose her contemplation strategy suboptimally. The individual's anticipation of possible over-thinking when choosing from a menu in the future leads to a violation of the monotonicity axiom that Ortoleva (2013) referred to as "thinking aversion."

[19]Although there are some minor differences in the assumptions imposed on the representations in this paper and Ergin and Saver (2010a), adapting the result to the current context is straightforward.

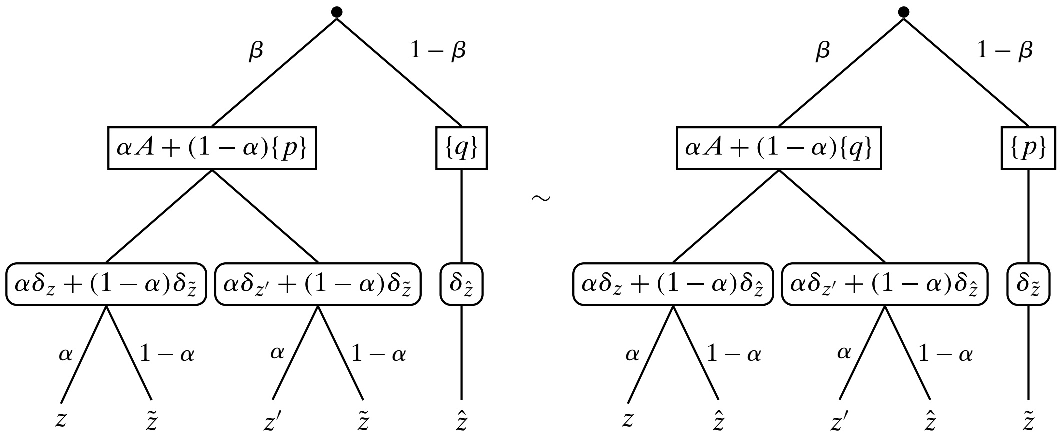

FIGURE 4. Reversibility of degenerate decisions when $A=\{\delta_z,\delta_{z'}\}$, $p=\delta_{\tilde z}$, and $q=\delta_{\hat z}$.

can be transformed into a consistent RFHA representation is as follows. In the CC representation, each contemplation strategy $\mathcal{G}$ leads to the random variable $\mathbb{E}_\pi[U|\mathcal{G}]$, which denotes the individual's ex post expected-utility function after acquiring signal $\mathcal{G}$. Thus, each contemplation strategy $\mathcal{G}$ can be associated with the distribution over ex post utility functions over $\triangle(Z)$ that it induces. Moreover, the law of iterated expectations implies that for any contemplation strategy $\mathcal{G}$, the ex ante expected value of the ex post utility function $\mathbb{E}_\pi[U|\mathcal{G}]$ must agree with the utility function prior to acquiring any information, $\mathbb{E}_\pi[U]$, which implies the consistency condition on the corresponding set of measures.

Given an RFHA representation $(\mathcal{M}, c)$, we will show that the following axiom captures consistency of the set of measures.

AXIOM 5 (Reversibility of degenerate decisions (RDD)). *For any $A\in\mathcal{A}$, $p,q\in\triangle(Z)$, and $\alpha\in[0,1]$,*

$$\beta\delta_{\alpha A+(1-\alpha)\{p\}}+(1-\beta)\delta_{\{q\}}\sim\beta\delta_{\alpha A+(1-\alpha)\{q\}}+(1-\beta)\delta_{\{p\}},$$

where $\beta=1/(2-\alpha)$.

We will call a choice out of a singleton menu a degenerate decision. To interpret Axiom 5, consider first the lottery $\beta\delta_{\alpha A+(1-\alpha)\{p\}}+(1-\beta)\delta_{\{q\}}$. Under this lottery, the individual makes a choice out of the menu $\alpha A+(1-\alpha)\{p\}$ with probability β, and makes a degenerate choice out of the menu $\{q\}$ with probability $1-\beta$. A choice out of the menu $\alpha A+(1-\alpha)\{p\}$ can be interpreted as a contingent plan, where initially in period 2 the individual determines a lottery out of A, and then her choice out of A is executed with probability α and the fixed lottery p is executed with the remaining $1-\alpha$ probability. The lottery $\beta\delta_{\alpha A+(1-\alpha)\{q\}}+(1-\beta)\delta_{\{p\}}$ has a similar interpretation with the roles of p and q reversed. Figure 4 illustrates these two lotteries for the case where $A=\{\delta_z,\delta_{z'}\}$, $p=\delta_{\tilde z}$, and $q=\delta_{\hat z}$.

If one interprets the individual's behavior as one of costly contemplation/subjective information acquisition, then her optimal contemplation strategy might change as the probability α that her choice out of A is executed changes, since her return to contemplation will be higher for higher values of α. However, since the probability that her choice out of A will be executed is the same in both $\alpha A+(1-\alpha)\{p\}$ and $\alpha A+(1-\alpha)\{q\}$, it is reasonable to expect that her contemplation strategy would be the same for both contingent planning problems. Still, she need not be indifferent between $\delta_{\alpha A+(1-\alpha)\{p\}}$ and $\delta_{\alpha A+(1-\alpha)\{q\}}$ depending on her preference between $\delta_{\{p\}}$ and $\delta_{\{q\}}$. Similarly, depending on her preference between $\delta_{\{p\}}$ and $\delta_{\{q\}}$, she need not be indifferent between the lotteries $\beta\delta_{\alpha A+(1-\alpha)\{p\}}+(1-\beta)\delta_{\{q\}}$ and $\beta\delta_{\alpha A+(1-\alpha)\{q\}}+(1-\beta)\delta_{\{p\}}$ if the probabilities of the paths leading to p and q, i.e., $\beta(1-\alpha)$ and $1-\beta$, are different. The RDD axiom requires the individual to be indifferent between these two lotteries when the probabilities of these paths are the same, i.e., when $\beta(1-\alpha)=1-\beta$ or, equivalently, $\beta=1/(2-\alpha)$. In the example illustrated in Figure 4, in both trees, the probabilities of the paths leading to $\tilde{z}$ and $\hat{z}$ are the same when $\beta=1/(2-\alpha)$.

We next present the main result of this section. Given an RFHA representation $(\mathcal{M},c)$, we show that RDD is equivalent to consistency of $(\mathcal{M},c)$.

Theorem 5. *Suppose that the preference $\succsim$ has an RFHA representation $(\mathcal{M},c)$. Then $(\mathcal{M},c)$ satisfies consistency if and only if $\succsim$ satisfies RDD.*

The following CC representation theorem is obtained from Theorems 2, 4, and 5.

Corollary 2. *The preference $\succsim$ has a CC representation if and only if it satisfies Axiom 1, PERU, RDD, and monotonicity.*

By Corollary 2, a preference with a CC representation satisfies PERU. However, it is immediate from the representation that such a preference always satisfies indifference to timing of resolution of uncertainty when restricted to temporal lotteries, i.e., for all $p,q\in\triangle(Z)$ and $\alpha\in(0,1)$,

$$\alpha\delta_{\{p\}}+(1-\alpha)\delta_{\{q\}}\sim\delta_{\{\alpha p+(1-\alpha)q\}}.\text{[20]}$$

Therefore, as suggested at the beginning of this section, an individual with CC preferences never has a strict PERU unless she has nondegenerate choices in period 2.

[20] This property can also be established directly as a consequence of RDD and first-stage independence. Fix any $p,q\in\triangle(Z)$ and $\alpha\in(0,1)$. Letting $\beta=1/(2-\alpha)$ and $A=\{p\}$, RDD implies

$$\beta\delta_{\{p\}}+(1-\beta)\delta_{\{q\}}\sim\beta\delta_{\{\alpha p+(1-\alpha)q\}}+(1-\beta)\delta_{\{p\}}.$$

Since $\beta=1/(2-\alpha)$ implies that $\beta=1-\beta+\alpha\beta$ and $1-\beta=(1-\alpha)\beta$, the left side of this expression is equal to $(1-\beta)\delta_{\{p\}}+\alpha\beta\delta_{\{p\}}+(1-\alpha)\beta\delta_{\{q\}}$. Hence,

$$\beta[\alpha\delta_{\{p\}}+(1-\alpha)\delta_{\{q\}}]+(1-\beta)\delta_{\{p\}}\sim\beta\delta_{\{\alpha p+(1-\alpha)q\}}+(1-\beta)\delta_{\{p\}},$$

which, by first-stage independence, implies $\alpha\delta_{\{p\}}+(1-\alpha)\delta_{\{q\}}\sim\delta_{\{\alpha p+(1-\alpha)q\}}$.

4.2 *Simple models of preference for flexibility and timing*

In this section, we describe a simple model that allows for both preference for flexibility and preference for early resolution of uncertainty, but does not allow the preference for early resolution of uncertainty to depend on the content of the menu per se. The approach we follow here parallels that of Kreps and Porteus (1978). They were able to incorporate preference for early resolution of uncertainty into a standard expected-utility model by taking a nonlinear transformation of second-stage expected utility before taking expectations with respect to first-stage uncertainty. Formally, period 2 choice in their model maximizes some expected-utility function v and, thus, menus in the second period are evaluated by $\max_{p\in A} v(p)$. This utility value is then transformed by some function ϕ to obtain the Bernoulli utility index for first-stage uncertainty:

$$V(A)=\phi\Big(\max_{p\in A} v(p)\Big). \tag{6}$$

Notice that $\alpha V(A)+(1-\alpha)V(B)\geq V(\alpha A+(1-\alpha)B)$ for all $A,B\in\mathcal{A}$ if and only if ϕ is convex. Therefore, for the first-stage expected-utility representation $\mathbb{E}_P[V]$, PERU corresponds to convexity of ϕ.

This approach of using a nonlinear transformation to alter preferences for timing can be applied to many models beyond standard expected utility. For example, to allow for preference for flexibility, suppose menus in the second period are evaluated by the DLR (2001) additive representation $\int_{\mathcal{U}} \max_{p\in A} u(p)\mu(du)$ for some measure μ on the set of expected-utility functions $\mathcal{U}$. As in Kreps and Porteus (1978), we can transform this utility value by a function ϕ to incorporate preferences for early resolution of uncertainty. This suggests the following representation, which includes the (two-stage) Kreps–Porteus representation as a special case.[21]

Definition 4. A *Kreps–Porteus–Dekel–Lipman–Rustichini* (KPDLR) representation is a pair (ϕ,μ), where μ is a finite Borel measure on $\mathcal{U}$, and $\phi:[a,b]\to\mathbb{R}$ is a Lipschitz continuous and strictly increasing function on the bounded interval $[a,b]=\{\int_{\mathcal{U}}\max_{p\in A} u(p)\mu(du): A\in\mathcal{A}\}$, such that $P\succsim Q$ if and only if $\mathbb{E}_P[V]\geq\mathbb{E}_Q[V]$, where $V:\mathcal{A}\to\mathbb{R}$ is defined by

$$V(A)=\phi\bigg(\int_{\mathcal{U}}\max_{p\in A} u(p)\mu(du)\bigg). \tag{7}$$

A *Kreps–Porteus representation* is a KPDLR representation where $\mu=\alpha\delta_u$ for some $u\in\mathcal{U}$ and $\alpha\geq 0$.[22]

While the KPDLR representation (and the Kreps–Porteus representation in particular) has the virtue of being relatively parsimonious, its drawback is that it places nontrivial restrictions on the possible preferences for timing. To illustrate, consider any two

[21] Kraus and Sagi (2006, Theorem 5.2) also studied a similar generalization of the DLR (2001) additive representation and the Kreps–Porteus representation for incomplete preferences.

[22] A KPDLR representation (ϕ,μ) in which $\mu=\alpha\delta_u$ for $\alpha\geq 0$ corresponds to the Kreps–Porteus formulation in (6) for the expected-utility function $v=\alpha u$.

menus A and B such that $V(A) = V(B)$. It then follows from (7) that $V(\alpha A + (1-\alpha)C) = V(\alpha B + (1-\alpha)C)$ for any other menu C. This implies that the preference for timing of resolution of uncertainty exhibited for the menus A and C must be the same as that exhibited for the menus B and C. In fact, the utility difference between early and late resolution of uncertainty must be the same in both cases:

$$\begin{aligned}\alpha V(A) + (1-\alpha)V(C) - V(\alpha A + (1-\alpha)C)& \\ = \alpha V(B) + (1-\alpha)V(C) - V(\alpha B + (1-\alpha)C)&.\end{aligned}$$

This shows that the preference for timing does not depend directly on the content of the menus, only on the resulting utility values.

In particular, if a nonsingleton menu A satisfies $V(A) = V(\{p\})$ for some lottery p, the preference for timing for two-stage lotteries involving A is the same as for lotteries where $\{p\}$ takes the place of A. This illustrates why in the Kreps–Porteus representation, the preference for timing for general two-stage lotteries is completely determined by the preference for timing for temporal lotteries (without period 2 choice). This feature of their model is in contrast to the costly contemplation representation from the previous section, where the individual is indifferent to timing of resolution of uncertainty when choosing among temporal lotteries, but may exhibit a strict PERU when she faces nondegenerate choices in period 2.

To better illustrate the connection with our general model, we now provide an axiomatic treatment of the KPDLR model and describe how it can be formulated as a special case of our RFHA representation. Since the value function in (7) is a monotone transformation of the DLR (2001) additive representation, it follows that for any menus A and B,

$$\delta_A \succsim \delta_B \quad \Leftrightarrow \quad \int_{\mathcal{U}} \max_{p \in A} u(p)\mu(du) \geq \int_{\mathcal{U}} \max_{p \in B} u(p)\mu(du).$$

Therefore, the KPDLR representation must satisfy the DLR (2001) axioms on the restricted domain of degenerate lotteries over menus. In fact, since V is determined up to a monotonic transformation by the ranking of lotteries δ_A for $A \in \mathcal{A}$, their axioms are also sufficient for the KPDLR representation (when combined with Axiom 1). Aside from weak order, continuity, and monotonicity, which were stated above, the key axiom for their representation is an independence axiom for menus. The following axiom is a translation of their axiom to our two-stage setting.[23]

Axiom 6 (Mixture independence). *For any $A, B, C \in \mathcal{A}$ and $\alpha \in (0,1)$,*

$$\delta_A \succ \delta_B \quad \Rightarrow \quad \delta_{\alpha A + (1-\alpha)C} \succ \delta_{\alpha B + (1-\alpha)C}.$$

[23]The restrictions on preferences for timing in the KPDLR representation are also easily established as direct implications of the axioms. By continuity and mixture independence, if $\delta_A \sim \delta_B$, then $\delta_{\alpha A + (1-\alpha)C} \sim \delta_{\alpha B + (1-\alpha)C}$ for any menu C. By continuity and first-stage independence, $\alpha\delta_A + (1-\alpha)\delta_C \sim \alpha\delta_B + (1-\alpha)\delta_C$. Thus, the preference for timing of resolution of uncertainty must be the same for A and C as for B and C.

To obtain the more specialized Kreps–Porteus representation, we need to strengthen the monotonicity axiom to ensure there is no strict preference for flexibility. This is accomplished by the following axiom from Kreps (1979), which guarantees that the individual is indifferent between any menu and its best singleton subset. Kreps and Porteus (1978) implicitly assume the same relationship between the individual's ranking of menus and alternatives.[24]

Axiom 7 (Strategic rationality). *For any $A, B \in \mathcal{A}$, $\delta_A \succsim \delta_B$ implies $\delta_A \sim \delta_{A\cup B}$.*

The following result formalizes the connection between these axioms and the KPDLR representation.

Theorem 6. *A. The preference $\succsim$ has a KPDLR representation if and only if it satisfies Axiom 1, mixture independence, and monotonicity.*

B. (Kreps and Porteus 1978) The preference $\succsim$ has a Kreps–Porteus representation if and only if it satisfies Axiom 1, mixture independence, and strategic rationality.[25]

C. If the preference $\succsim$ has the KPDLR representation (ϕ, μ), then $\succsim$ satisfies PERU if and only if ϕ is convex.

Note that KPDLR preferences need not be a subset of HA preferences, since they may violate PERU. However, Theorem 6.C shows that PERU is satisfied whenever ϕ is convex. The following theorem describes how this subclass of KPDLR representations can be expressed as special cases of our RFHA representation.

Theorem 7. *Let $V : \mathcal{A} \to \mathbb{R}$ and let μ be a nonzero finite Borel measure on $\mathcal{U}$. Then the following statements are equivalent:*

(i) There exists a KPDLR representation (ϕ, μ) with convex ϕ such that V is given by (7).

(ii) There exists an RFHA representation $(\mathcal{M}, c)$ such that V is given by (4), where

(a) $\mathcal{M} \subset \{\lambda\mu : \lambda \in \mathbb{R}_+\}$

(b) 0 is not an isolated point of $\mathcal{M}$ and if $0 \in \mathcal{M}$, then

$$\lim_{\lambda \searrow 0 :\, \lambda\mu \in \mathcal{M}} \frac{c(\lambda\mu) - c(0)}{\lambda} = \min_{A \in \mathcal{A}} \int_{\mathcal{U}} \max_{p \in A} u(p)\mu(du).$$

[24] To be precise, Kreps and Porteus (1978) considered both a period 1 preference $\succsim$ over first-stage lotteries in $\triangle(\mathcal{A})$ and a period 2 preference $\succsim_2$ over second-stage lotteries in $\triangle(Z)$. It is easy to show that imposing their temporal consistency axiom (Axiom 3.1 in their paper) on this pair of preferences $(\succsim, \succsim_2)$ implies that the period 1 preference $\succsim$ satisfies strategic rationality. Conversely, if the period 1 preference $\succsim$ satisfies strategic rationality along with continuity, then there exists *some* period 2 preference $\succsim_2$ such that the pair $(\succsim, \succsim_2)$ satisfies their temporal consistency axiom. Moreover, in this case, the period 1 preference $\succsim$ satisfies our mixture independence axiom if and only if this period 2 preference $\succsim_2$ satisfies the substitution axiom of Kreps and Porteus (1978, Axiom 2.3).

[25] Kreps and Porteus (1978) only required that the transformation ϕ be continuous. We additionally require Lipschitz continuity of ϕ since we impose the L-continuity axiom throughout the paper.

Theorem 7 shows that for any KPDLR representation satisfying PERU, the hidden actions in the corresponding RFHA representation can be indexed by a real number λ. In particular, condition (ii)(a) shows that every hidden action is a scalar multiple of a fixed measure μ. Condition (ii)(b) is merely a technical regularity condition on the derivative of the cost function c at 0 that ensures that ϕ is strictly increasing.

The form of the RFHA representation in condition (ii) suggests the following interpretation: The distribution of possible tastes (determined by μ) is the same for every hidden action, and changing the action simply changes the magnitude of the ex post utilities by a common scalar multiple λ. The optimal action for a given menu A is, therefore, determined entirely by the value $\int_{\mathcal{U}} \max_{p \in A} u(p)\mu(du)$ and the shape of the cost function c. This implies that if this integral expression takes the same value for any two menus A and B, the optimal hidden action must be the same for both menus. Consequently, the preference for timing in any two-stage lottery involving A must be the same if B takes the place of A in the lottery. Thus, condition (ii) corresponds to the same restrictions on preferences for timing that were described already for the KPDLR representation, but expresses them in a different way: Condition (ii) restricts the possible preferences for timing by placing strong restrictions on the complementarities between menus and hidden actions.

As noted above, one benefit of the KPDLR representation is that it is relatively parsimonious. Taking this perspective, one can think of Theorem 7 as describing the instances in which the KPDLR representation can be used as a reduced-form representation for a hidden action model. However, since condition (ii) is rather restrictive, one implication of this result is that the KPDLR representation will only be appropriate in a fairly limited set of circumstances.[26] Since the Kreps–Porteus representation is a special case—where $\mu = \alpha\delta_u$ and the corresponding hidden action representation exhibits no uncertainty about the ex post preference ranking—this argument applies to that model a fortiori.

On a final note, Theorem 7 generalizes several results from Kreps and Porteus (1979), who considered a class of hidden action representations and determined the conditions on the representation under which the resulting preference satisfies the axioms of Kreps and Porteus (1978). Specifically, Propositions 5 and 6 in Kreps and Porteus (1979) show that a hidden action representation corresponds to a Kreps–Porteus preference if and only if it takes a functional form that is essentially equivalent to the one described by condition (ii) for a measure of the form $\mu = \alpha\delta_u$. Thus, their results follow when Theorem 7 is applied to measures taking the Kreps–Porteus form described in Definition 4.

5. Discussion of second-period stochastic choice

Throughout our analysis, we have focused attention on the individual's first-period choice of lotteries over menus. However, the interpretation of the HA representation

[26] It is well known that independence will, in general, be violated if the individual takes a payoff-relevant action prior to the resolution of uncertainty; for instance, see Markowitz (1959, Chapters 10–11), Mossin (1969), and Spence and Zeckhauser (1972). Theorem 7 characterizes precisely those special cases in which independence is not violated.

suggests a specific probabilistic structure for second-period choice: Intuitively, the probability an alternative $p \in A$ will be selected from the menu A is equal to the probability according to $\pi(\cdot;\theta)$ of the states ω in which $p \in \arg\max_{q\in A} U(q,\omega;\theta)$, where θ is the optimal hidden action for the menu A.[27]

In this section, we provide a brief informal discussion of how second-period stochastic choice might be incorporated into the model. While a detailed formal analysis of second-period choice is beyond the scope of the current paper, we hope the observations in this section serve as a useful starting point for future research in this direction. This section is divided into two parts: In Section 5.1, we explain how the identification of the HA representation can be improved using second-period choice. In Sections 5.2 and 5.3, we outline some necessary conditions for second-period choice to be consistent with the model. Our discussion will be a bit heuristic at times, and we will omit some technical details and precise definitions. The interested reader can consult the papers cited herein for a more formal description.

5.1 *Using second-period choice to identify beliefs*

As discussed in the Introduction and Section 3.2, it is impossible to separately identify state-dependent utility and beliefs in the HA representation, which prompted us to combine cardinality of utility and probability into a single (nonprobability) measure in our RFHA representation. However, by also using second-period choice frequencies from menus, it may be possible to uniquely pin down probabilities in the HA representation, thereby overcoming the nonuniqueness issue and alleviating the need to use the reduced-form representation.

To illustrate this approach, consider first the special case of the HA representation where Θ is a singleton and V therefore takes the form of the additive representation from DLR (2001):

$$V(A) = \int_\Omega \max_{p\in A} U(p,\omega)\pi(d\omega).$$

Ahn and Sarver (2013) showed that for this additive representation, preferences over menus together with stochastic choice from menus provide sufficient information to identify both the state-dependent utility function $U(p,\omega)$ and the probability measure π.[28] This result is formally based on a combination of the uniqueness properties of the DLR (2001) representation for preferences over menus with the uniqueness properties of the Gul and Pesendorfer (2006) random expected-utility representation for stochastic choice. The intuition for this uniqueness can be understood using a simple example. Fix any two ordinally distinct expected-utility functions u_1 and u_2, and consider the following two possible additive representations on the state space $\Omega = \{\omega_1,\omega_2\}$:

[27] This heuristic description omits several important technical considerations, such as the possibility that p is one of multiple maximizers in a particular state. These issues are addressed in detail in Gul and Pesendorfer (2006); see also Ahn and Sarver (2013).

[28] After imposing a nonredundancy condition (the utility function does not induce the same ex post preference over lotteries for two distinct states), the utility function in their model is unique up to relabeling of states and a state-independent affine transformation. The probability measure is unique up to the same relabeling of states.

	ω_1	ω_2
$U(p,\omega)$	$u_1(p)$	$u_2(p)$
$\pi(\omega)$	1/2	1/2

	ω_1	ω_2
$\hat{U}(p,\omega)$	$(3/2)u_1(p)$	$(3/4)u_2(p)$
$\hat{\pi}(\omega)$	1/3	2/3

Notice that (U, π) and $(\hat{U}, \hat{\pi})$ both yield the same value function V for menus, and therefore cannot be distinguished using first-period choice. However, these two representations are easy to differentiate if we also observe second-period choice. Take any p and q such that $u_1(p) > u_1(q)$ and $u_2(q) > u_2(p)$. Then the probability that p is chosen from the menu $A = \{p, q\}$ is precisely the probability of the state ω_1; likewise, the probability that q is chosen from this menu is the probability of the state ω_2. Thus the choice frequencies from this menu pin down the probability measure in the representation, which, together with first-period choice, also pins down the utility function (up to a state-independent affine transformation).

Extending this approach for the case of nontrivial hidden actions is straightforward in some cases, but more difficult in others. For example, in the nonreduced-form version of the KPDLR representation described in Table 1, the probability measure π is independent of θ and $U(p, \omega; \theta) = f(p, \omega)g(\theta)$. Since the probability of each ex post preference is independent of the hidden action, second-period choice is again sufficient to pin down this probability measure uniquely. Then, since our uniqueness result for the RFHA representation (Theorem 3) implies the product of utility and probability is uniquely identified for every hidden action, we can back out the function $U(p, \omega; \theta)$.

However, the general case where the probability measure π varies with the hidden action is less straightforward. For example, suppose θ is the optimal hidden action for some menu A, and for this particular hidden action, suppose $U(p, \omega; \theta)$ and $\pi(\omega; \theta)$ take the same values as $U(p, \omega)$ and $\pi(\omega)$ in the left table of the example above. How do we differentiate this representation from alternatives (such as $\hat{U}(p, \omega)$ and $\hat{\pi}(\omega)$ in the right table above) using choice data? As in that example, we would need to find some alternatives p and q such that p is preferred in the first state and q is preferred in the second. However, the optimal hidden action for the menu $\{p, q\}$ (or potentially any menu containing p and q) may be some other θ', not θ. Consequently, choice frequencies between p and q may not be useful for identifying the probability measure $\pi(\cdot; \theta)$ for the hidden action in question. An important open question is therefore whether—perhaps by perturbing slightly some of the alternatives in the original menu A in the direction of p and some in the direction of q (e.g., using convex combinations)—it is still possible to pin down the probability measure $\pi(\omega; \theta)$.

5.2 *Structure of second-period choice*

To understand the implications of our model for second-period stochastic choice, it is useful to first contrast with the well established literature on random utility models. As noted above, if the individual chooses the hidden action θ when faced with a menu A, the interpretation of second-period choice in the HA representation is the following: The probability that an alternative $p \in A$ will be selected is equal to the probability according to $\pi(\cdot; \theta)$ of the states ω in which $p \in \arg\max_{q \in A} U(q, \omega; \theta)$ (ignoring the possibility of indifferences between alternatives in A for expositional simplicity). This choice procedure for fixed θ is referred to as a random utility model. Random utility encompasses a number of well known models of stochastic choice, including the Luce (1959) model and many econometric models such as logit and probit.

We will see that some special cases of our HA representation will lead to second-period choice that is consistent with random utility maximization;[29] however, in general, since the optimal θ can vary with the menu, we can observe behavior that is inconsistent with the random utility model. To illustrate, we will first summarize the conditions on stochastic choice that characterize random utility, and then explain why these must be relaxed to accommodate second-period choice in the HA model.

In settings with deterministic alternatives, behavioral conditions for random utility models were provided by Falmagne (1978), McFadden and Richter (1990), and Clark (1996). In our domain of choice from lotteries, a random *expected-utility* model was proposed by Gul and Pesendorfer (2006). They showed that this model is equivalent to a continuity axiom plus the following three easily interpreted axioms on stochastic choice behavior:

- *Monotonicity*: If $p \in A \subset B$, then the probability of choosing p from B is no greater than the probability of choosing p from A.[30]
- *Linearity*: The probability of choosing $\alpha p + (1-\alpha)q$ from the menu $\alpha A + (1-\alpha)\{q\}$ is the same as the probability of choosing p from A.
- *Extreme*: Only extreme points of a menu A are chosen with positive probability.

In the special cases of our HA representation where the probability distribution over ex post preferences is independent of the hidden action, second-period choice will satisfy these axioms. This corresponds to the last entry in Table 1, where π is independent of θ and $U(p, \omega; \theta) = f(p, \omega)g(\omega, \theta)$ (for a function $g > 0$). This form of the utility function implies that although the hidden action can affect the relative cardinality of utility functions across states through the term $g(\omega, \theta)$, it does not impact the preferences over lotteries for any state ω. Consequently, the distribution of second-period choices is unaffected by the hidden action. This special case of the HA model includes the KPDLR representation and, therefore, also the DLR (2001) representation (where Θ is a singleton).

The HA representation will always satisfy the extreme axiom since any strict maximizer of the ex post expected-utility function $U(p, \omega; \theta)$ will be an extreme point of the choice set.[31] However, both the monotonicity and linearity axioms will, in general, be violated. To illustrate these violations concretely, consider the following example of the costly contemplation representation. Suppose menus consist of lotteries

[29]This obviously includes the special case of the DLR (2001) model (i.e., Θ is a singleton) since the representation for second-period choice in this case is precisely a random expected-utility function.

[30]This condition is a probabilistic version of Sen's α for deterministic choice, which states $z \in A \subset B$ and $z \in \mathcal{C}(B)$ implies $z \in \mathcal{C}(A)$.

[31]In the case of multiple maximizers, we would need to incorporate a tie-breaking rule to ensure that second-period choice is well defined. If we adopt the common practice of using expected-utility functions also to break ties (see Gul and Pesendorfer 2006 or Ahn and Sarver 2013), then in this case as well only extreme points will be selected.

over three deterministic alternatives, $Z = \{z_1, z_2, z_3\}$. Although each of these alternatives is (objectively) deterministic, suppose z_1 and z_2 are subjectively risky in the sense that the individual may like or dislike each a great deal, and, moreover, she must exert mental effort to determine which she will like. Alternative z_3 is a subjectively safe choice. Formally, suppose there are two equally likely subjective states, $\Omega = \{\omega_1, \omega_2\}$ and $\pi(\omega_1) = \pi(\omega_2) = \frac{1}{2}$, and the Bernoulli utilities for the three alternatives in the two states are given by the vectors

$$\bar{U}(\omega_1) = (5, -5, 1) \quad \text{and} \quad \bar{U}(\omega_2) = (-5, 5, 1).$$

As summarized in Table 1, in the costly contemplation representation, the hidden actions are signals about the subjective state, and the utility function conditional on the hidden action is the conditional expectation of $\bar{U}$ given the signal. Suppose θ_1 corresponds to learning nothing about the state, and θ_2 corresponds to full information about the state. Then, for any lottery $p \in \triangle(Z)$ and state $\omega \in \Omega$,

$$U(p, \omega; \theta_1) = \mathbb{E}[\bar{U}] \cdot p = (0, 0, 1) \cdot p \quad \text{and} \quad U(p, \omega; \theta_2) = \bar{U}(\omega) \cdot p.$$

Finally, suppose the cost of no information is $c(\theta_1) = 0$ and the cost of full information is $c(\theta_2) = 3$ (the conclusions on this example hold for any $2 < c(\theta_2) < 4$).

For these parameters, it is easy to verify that the optimal hidden action for the menu $\{\delta_{z_1}, \delta_{z_3}\}$ is θ_1 (no contemplation) and the optimal hidden action for the menu $\{\delta_{z_1}, \delta_{z_2}, \delta_{z_3}\}$ is θ_2 (contemplation). Intuitively, having both z_1 and z_2 in the same menu creates sufficient option value in discovering the true state to overcome the cost of contemplation. However, this implies that δ_{z_1} is chosen with probability zero from $\{\delta_{z_1}, \delta_{z_3}\}$ but with probability $\frac{1}{2}$ from $\{\delta_{z_1}, \delta_{z_2}, \delta_{z_3}\}$, violating the monotonicity axiom. Likewise, the independence axiom is violated since the probability of choosing $\alpha\delta_{z_1} + (1-\alpha)\delta_{z_3}$ from the menu $\alpha\{\delta_{z_1}, \delta_{z_2}, \delta_{z_3}\} + (1-\alpha)\{\delta_{z_3}\}$ is $\frac{1}{2}$ for $\alpha \approx 1$ and drops to zero for $\alpha \approx 0$ (since the optimal hidden action switches from θ_2 to θ_1 for α sufficiently small). While this example shows that the monotonicity and independence axioms for stochastic choice are too restrictive for the HA representation, it is an open question to determine what (if any) relaxation of these axioms is implied by the HA model.

5.3 *Linking first- and second-period choice*

In addition to the structure the HA representation imposes on second-period choice in isolation, it also imposes some joint restrictions on first- and second-period choice. In related work, Ahn and Sarver (2013) found two conditions that allow preferences over menus and stochastic choice from menus to be represented by an additive representation from DLR (2001) and random expected-utility representation from Gul and Pesendorfer (2006), respectively, where the representations for both stages of choice share the same subjective state space, beliefs, and state-dependent utility function. Their first linking axiom states that if the individual strictly prefers to add a lottery p to the menu A in the first period, then she must choose p with positive probability from the menu

$A \cup \{p\}$ in the second period. The second axiom is roughly the converse: After controlling for the possibility of ties, if p is chosen with positive probability from $A \cup \{p\}$ in the second period, then adding p to the menu A is strictly preferred in the first period.

It is easy to see that both of these axioms are also necessary conditions whenever first- and second-period choice can be represented using a single HA representation. Therefore, after determining the appropriate axioms on second-period choice (in isolation) that are consistent with the HA model, the last remaining question is whether these linking axioms are also sufficient to represent first- and second-period choices that are each consistent with the HA representation using the *same* HA representation.

Appendix A: Mathematical preliminaries

In this section, we present some general mathematical results that will be used to prove our representation and uniqueness theorems. Our main results will center around a classic duality relationship from convex analysis. Throughout this section, let X be a real Banach space and let X^* denote the space of all continuous linear functionals on X.

Definition 5. Suppose $C \subset X$. A function $f : C \to \mathbb{R}$ is said to be *Lipschitz continuous* if there is some real number K such that $|f(x) - f(y)| \leq K\|x - y\|$ for every $x, y \in C$. The number K is called a *Lipschitz constant* of f.

We now introduce the standard definition of the subdifferential of a function.

Definition 6. Suppose $C \subset X$ and $f : C \to \mathbb{R}$. For $x \in C$, the *subdifferential* of f at x is defined to be

$$\partial f(x) = \{x^* \in X^* : \langle y - x, x^* \rangle \leq f(y) - f(x) \text{ for all } y \in C\}.$$

The subdifferential is useful for the approximation of convex functions by affine functions. It is straightforward to show that $x^* \in \partial f(x)$ if and only if the affine function $h : X \to \mathbb{R}$ defined by $h(y) = f(x) + \langle y - x, x^* \rangle$ satisfies $h \leq f$ and $h(x) = f(x)$. It should also be noted that when X is infinite-dimensional, it is possible to have $\partial f(x) = \varnothing$ for some $x \in C$, even if f is convex. However, the following result shows that a Lipschitz continuous and convex function always has a nonempty subdifferential.

Lemma 3 (Gale 1967). *Suppose C is a convex subset of a Banach space X. If $f : C \to \mathbb{R}$ is Lipschitz continuous and convex, then $\partial f(x) \neq \varnothing$ for all $x \in C$.*

We now introduce the definition of the conjugate of a function.

Definition 7. Suppose $C \subset X$ and $f : C \to \mathbb{R}$. The *conjugate* (or *Fenchel conjugate*) of f is the function $f^* : X^* \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(x^*) = \sup_{x \in C}[\langle x, x^* \rangle - f(x)].$$

There is an important duality between f and f^*. Lemma 4 summarizes certain properties of f^* that are useful in establishing this duality.[32] The proof is standard and can be found, for example, in the Supplemental Material of Ergin and Saver (2010a).

LEMMA 4. *Suppose $C \subset X$ and $f : C \to \mathbb{R}$. Then the following statements hold:*

(i) We have that f^* is lower semicontinuous in the weak topology.*

(ii) We have $f(x) \geq \langle x, x^* \rangle - f^*(x^*)$ for all $x \in C$ and $x^* \in X^*$.

(iii) We have $f(x) = \langle x, x^* \rangle - f^*(x^*)$ if and only if $x^* \in \partial f(x)$.

Suppose that $C \subset X$ is convex and $f : C \to \mathbb{R}$ is Lipschitz continuous and convex. As noted above, this implies that $\partial f(x) \neq \varnothing$ for all $x \in C$. Therefore, by parts (ii) and (iii) of Lemma 4, we have

$$f(x) = \max_{x^* \in X^*} [\langle x, x^* \rangle - f^*(x^*)] \tag{8}$$

for all $x \in C$.[33] To establish the existence of a minimal set of measures in the proof of Theorem 2, it is useful to establish that under certain assumptions, there is a minimal compact subset of X^* for which (8) holds. Let C_f denote the set of all $x \in C$ for which the subdifferential of f at x is a singleton:

$$C_f = \{x \in C : \partial f(x) \text{ is a singleton}\}. \tag{9}$$

Let $\mathcal{N}_f$ denote the set of functionals contained in the subdifferential of f at some $x \in C_f$:

$$\mathcal{N}_f = \{x^* \in X^* : x^* \in \partial f(x) \text{ for some } x \in C_f\}. \tag{10}$$

Finally, let $\mathcal{M}_f$ denote the closure of $\mathcal{N}_f$ in the weak* topology:

$$\mathcal{M}_f = \bar{\mathcal{N}}_f. \tag{11}$$

Before stating our first main result, recall that the *affine hull* of a set $C \subset X$, denoted $\text{aff}(C)$, is defined to be the smallest affine subspace of X that contains C. Also, a set $C \subset X$ is said to be a *Baire space* if every countable intersection of dense open subsets of C is dense.

THEOREM 8 (Ergin and Sarver 2010b). *Suppose (i) X is a separable Banach space, (ii) C is a convex subset of X that is a Baire space (when endowed with the relative topology) such that* $\text{aff}(C)$ *is dense in X,*[34] *and (iii) $f : C \to \mathbb{R}$ is Lipschitz continuous and convex.*

[32] For a complete discussion of the relationship between f and f^*, see Ekeland and Turnbull (1983) or Holmes (1975). A finite-dimensional treatment can be found in Rockafellar (1970).

[33] This is a slight variation of the classic Fenchel–Moreau theorem. The standard version of this theorem states that if $f : X \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous and convex, then $f(x) = f^{**}(x) \equiv \sup_{x^* \in X^*} [\langle x, x^* \rangle - f^*(x^*)]$. See, e.g., Proposition 1 in Ekeland and Turnbull (1983, p. 97).

[34] In particular, if C is closed, then by the Baire category theorem, C is a Baire space. Also, note that if C contains the origin, then the affine hull of C is equal to the span of C.

Then $\mathcal{M}_f$ is weak compact, and for any weak* compact $\mathcal{M} \subset X^*$,*

$$\mathcal{M}_f \subset \mathcal{M} \quad \Leftrightarrow \quad f(x) = \max_{x^* \in \mathcal{M}} [\langle x, x^* \rangle - f^*(x^*)] \quad \forall x \in C.$$

The intuition for Theorem 8 is fairly simple. We already know from Lemma 4 that for any $x \in C_f$, $f(x) = \max_{x^* \in \mathcal{N}_f} [\langle x, x^* \rangle - f^*(x^*)]$. Ergin and Sarver (2010b) show that under the assumptions of Theorem 8, C_f is dense in C. Therefore, it can be shown that for any $x \in C$,

$$f(x) = \max_{x^* \in \mathcal{M}_f} [\langle x, x^* \rangle - f^*(x^*)].$$

In addition, if $\mathcal{M}$ is a weak* compact subset of X^* and $\mathcal{M}_f$ is not a subset of $\mathcal{M}$, then there exists $x^* \in \mathcal{N}_f$ such that $x^* \notin \mathcal{M}$. That is, there exists $x \in C_f$ such that $\partial f(x) = \{x^*\}$ and $x^* \notin \mathcal{M}$. Therefore, Lemma 4 implies $f(x) > \max_{x^* \in \mathcal{M}} [\langle x, x^* \rangle - f^*(x^*)]$.

In the proof of Theorem 2, we will construct an RFHA representation in which $\mathcal{M}_f$, for a certain function f, is the set of measures. We will then use the following result to establish that monotonicity leads to a positive set of measures. For this next result, assume that X is a Banach lattice.[35] Let $X_+ = \{x \in X : x \geq 0\}$ denote the *positive cone* of X. A function $f : C \to \mathbb{R}$ on a subset C of X is *monotone* if $f(x) \geq f(y)$ whenever $x, y \in C$ are such that $x \geq y$. A continuous linear functional $x^* \in X^*$ is *positive* if $\langle x, x^* \rangle \geq 0$ for all $x \in X_+$.

Theorem 9 (Ergin and Saver 2010a, Supplemental Material). *Suppose C is a convex subset of a Banach lattice X, such that at least one of the following conditions holds:*

(i) We have $x \vee x' \in C$ for any $x, x' \in C$.

(ii) We have $x \wedge x' \in C$ for any $x, x' \in C$.

Let $f : C \to \mathbb{R}$ be Lipschitz continuous, convex, and monotone. Then the functionals in $\mathcal{M}_f$ are positive.

Finally, the following result will be used in the proof of Theorem 3 to establish the uniqueness of the RFHA representation.

Theorem 10 (Ergin and Saver 2010a, Supplemental Material). *Suppose X is a Banach space and C is a convex subset of X. Let $\mathcal{M}$ be a weak* compact subset of X^* and let $c : \mathcal{M} \to \mathbb{R}$ be weak* lower semicontinuous. Define $f : C \to \mathbb{R}$ by*

$$f(x) = \max_{x^* \in \mathcal{M}} [\langle x, x^* \rangle - c(x^*)].$$

Then the following statements hold:

(i) The function f is Lipschitz continuous and convex.

[35] See Aliprantis and Border (1999, p. 302) for a definition of Banach lattices.

(ii) *For all $x \in C$, there exists $x^* \in \partial f(x)$ such that $x^* \in \mathcal{M}$ and $f^*(x^*) = c(x^*)$. In particular, this implies $\mathcal{N}_f \subset \mathcal{M}$, $\mathcal{M}_f \subset \mathcal{M}$, and $f^*(x^*) = c(x^*)$ for all $x^* \in \mathcal{N}_f$.*

(iii) *If C is also compact (in the norm topology), then $f^*(x^*) = c(x^*)$ for all $x^* \in \mathcal{M}_f$.*

Appendix B: Proof of Theorem 2 and some extensions

In this section, we prove Theorem 2. As an intermediate step in the construction, we first prove a general representation theorem for preferences that may violate monotonicity and we subsequently establish Theorems 2 as a special case. Since this general representation may be of independent interest, we define it below and state the representation result explicitly. The following generalization of the RFHA representation modifies Definition 2 by allowing for signed measures.

Definition 8. A *signed RFHA representation* is a pair $(\mathcal{M}, c)$ consisting of a compact set of finite signed Borel measures $\mathcal{M}$ on $\mathcal{U}$ and a lower semicontinuous function $c : \mathcal{M} \to \mathbb{R}$ such that the following statements hold:

(i) We have $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$, where $V : \mathcal{A} \to \mathbb{R}$ is defined by (4).

(ii) The set $\mathcal{M}$ is *minimal*: For any compact proper subset $\mathcal{M}'$ of $\mathcal{M}$, the function V' obtained by replacing $\mathcal{M}$ with $\mathcal{M}'$ in (4) is different from V.

It has been shown that an individual's preferences may violate monotonicity, referred to as a *preference for commitment*, due to psychological features such as regret and temptation (see, e.g., Sarver 2008, Gul and Pesendorfer 2001, and DLR 2009). Therefore, the signed HA representation may be a useful starting point for incorporating regret and temptation into our model of temporal preferences. However, we leave the study of specific violations of monotonicity that correspond to these phenomena within our model as a subject for future research.

In this section, we also consider another variation of the HA representation. Throughout the main text of the paper, our focus was on preference for early resolution of uncertainty. However, using the same techniques that will establish the HA representation, it is also possible to prove a counterpart to the HA representation result for preferences for late resolution of uncertainty. Since this variation is also of interest for future research and comes at little cost in terms of additional work, we state it formally. Analogous to our definition of preference for early resolution of uncertainty, a preference for late resolution of uncertainty is defined as follows.

Axiom 8 (Preference for late resolution of uncertainty (PLRU)). *For any $A, B \in \mathcal{A}$ and $\alpha \in (0,1)$,*

$$\delta_{\alpha A+(1-\alpha)B} \succsim \alpha\delta_A + (1-\alpha)\delta_B.$$

The following definition is a dual of the HA representation corresponding to PLRU. This representation involves a minimization instead of a maximization and, therefore, has a fundamentally different interpretation. Instead of thinking of the hidden action in this representation as being chosen by the individual, we interpret it as a malevolent nature acting adversely against the individual.[36] As with the HA representation, we will also consider a signed version of this dual representation.

Definition 9. A *malevolent nature (MN) representation* (*signed MN representation*) is a pair $(\mathcal{M}, c)$ consisting of a compact set of finite (signed) Borel measures $\mathcal{M}$ on $\mathcal{U}$ and a lower semicontinuous function $c : \mathcal{M} \to \mathbb{R}$ such that the following statements hold:

(i) We have $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$, where $V : \mathcal{A} \to \mathbb{R}$ is defined by

$$V(A) = \min_{\mu \in \mathcal{M}} \left(\int_{\mathcal{U}} \max_{p \in A} u(p) \mu(du) + c(\mu) \right). \tag{12}$$

(ii) The set $\mathcal{M}$ is *minimal*: For any compact proper subset $\mathcal{M}'$ of $\mathcal{M}$, the function V' obtained by replacing $\mathcal{M}$ with $\mathcal{M}'$ in (12) is different from V.

As noted in Section 3.2, our MN representation is related to a representation in Epstein et al. (2007) in the setting of menus of lotteries. The functional form of the representation in Epstein et al. (2007, Theorem 1) is a special case of the MN representation where $\mathcal{M}$ is a convex set of probability measures and the cost function is identically zero. However, there are two main distinctions between our models: First, our framework allows objective uncertainty to resolve in multiple stages, and we are, therefore, able to explicitly model the preferences for timing associated with the MN representation. The second and most significant difference is that they require a normalization on the state-dependent utility functions: they permit only one possible utility function for each ex post preference. Although such a normalization is without loss of generality for our MN representation due to the use of nonprobability measures (see Lemma 1), for representations that require probability measures, this requirement places nontrivial additional restrictions on the preference. This manifests in the Epstein et al. (2007) model as two auxiliary axioms needed to obtain their representation.[37]

We will say a pair $(\mathcal{M}, c)$ is a *signed representation* if it is a signed RFHA or a signed MN representation. With the definitions now established, we state our general representation theorem.

Theorem 11. *A. The preference $\succsim$ has a signed RFHA (MN) representation if and only if it satisfies Axiom 1, IR, and PERU (PLRU).*

B. The preference $\succsim$ has a RFHA (MN) representation if and only if it satisfies Axiom 1, IR, PERU (PLRU), and monotonicity.

[36]While we do not suggest that there literally exists a malevolent nature, it is a useful way to interpret a pessimistic attitude on the part of the decision-maker. See, for example, Maccheroni et al. (2006) for a related discussion in the context of ambiguity aversion.

[37]We are referring to the *Worst* and *Certainty Independence* axioms used in their Theorem 1. Epstein et al. (2007, p. 363) acknowledge that these two axioms are "excess baggage."

Theorem 11.A characterizes the signed RFHA and MN representations, and Theorem 11.B adds the assumption of monotonicity to obtain the (unsigned) RFHA and MN representations. Lemma 5 below shows that weak order, continuity, PERU, and monotonicity imply indifference to randomization (IR). Therefore, the RFHA representation part of Theorem 11.B is equivalent to Theorem 2.

LEMMA 5. *If a preference $\succsim$ satisfies weak order, continuity, PERU, and monotonicity, then it also satisfies IR.*

PROOF. The proof is similar to the proof of Lemma 2 in Ergin and Saver (2010a). We include it here for completeness. Let $A \in \mathcal{A}$. Monotonicity implies that $\delta_{\mathrm{co}(A)} \succsim \delta_A$; hence, we only need to prove that $\delta_A \succsim \delta_{\mathrm{co}(A)}$. Let us inductively define a sequence of sets via $A_0 = A$ and $A_k = \frac{1}{2}A_{k-1} + \frac{1}{2}A_{k-1}$ for $k \geq 1$. PERU implies that

$$\delta_{A_{k-1}} = \tfrac{1}{2}\delta_{A_{k-1}} + \tfrac{1}{2}\delta_{A_{k-1}} \succsim \delta_{(1/2)A_{k-1}+(1/2)A_{k-1}} = \delta_{A_k},$$

so, by transitivity, $\delta_A \succsim \delta_{A_k}$ for any k. It is straightforward to verify that $d_h(A_k, \mathrm{co}(A)) \to 0$, which implies that $\delta_{A_k} \to \delta_{\mathrm{co}(A)}$ in weak* topology. By continuity, we have $\delta_A \succsim \delta_{\mathrm{co}(A)}$. □

The remainder of this section is devoted to the proof of Theorem 11. Note that $\mathcal{A}$ is a compact metric space since $\triangle(Z)$ is a compact metric space (see, e.g., Munkres 2000, pp. 280–281, or Theorem 1.8.3 in Schneider 1993, p. 49). We begin by showing that weak order, continuity, and first-stage independence imply that $\succsim$ has an expected-utility representation.

LEMMA 6. *A preference $\succsim$ over $\triangle(\mathcal{A})$ satisfies weak order, continuity, and first-stage independence if and only if there exists a continuous function $V : \mathcal{A} \to \mathbb{R}$ such that $\succsim$ is represented by $\mathbb{E}_P[V]$. Furthermore, if $V : \mathcal{A} \to \mathbb{R}$ and $V' : \mathcal{A} \to \mathbb{R}$ are continuous functions such that $\mathbb{E}_P[V]$ and $\mathbb{E}_P[V']$ represent the same preference over $\triangle(\mathcal{A})$, then there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $V' = \alpha V + \beta$.*

Lemma 6 is a standard result. For example, it is asserted without proof in Corollary 5.22 of Kreps (1988). It is also a consequence of Theorem 10.1 of Fishburn (1970) together with some simple arguments to establish continuity of V from the continuity axiom.

Let $\mathcal{A}^c \subset \mathcal{A}$ denote the collection of all convex menus. It is a standard exercise to show that $\mathcal{A}^c$ is a closed subset of $\mathcal{A}$ and, hence, $\mathcal{A}^c$ is also compact (see Theorem 1.8.5 in Schneider 1993, p. 50). Our strategy for proving the sufficiency of the axioms will be to show that the function V described in Lemma 6 satisfies the RFHA (MN) formula on $\mathcal{A}^c$. Using the IR axiom, it will then be straightforward to show that V satisfies the RFHA (MN) formula on all of $\mathcal{A}$.

The following lemma shows the implications of our other axioms.

LEMMA 7. *Suppose that $V : \mathcal{A} \to \mathbb{R}$ is a continuous function such that $\mathbb{E}_P[V]$ represents the preference $\succsim$ over $\triangle(\mathcal{A})$. Then the following claims hold:*

(i) If $\succsim$ satisfies L-continuity, then V is Lipschitz continuous on $\mathcal{A}^c$, i.e., there exists $K \geq 0$ such that $|V(A) - V(B)| \leq K d_h(A, B)$ for any $A, B \in \mathcal{A}^c$.[38]

(ii) If V is Lipschitz continuous (on $\mathcal{A}$), then $\succsim$ satisfies L-continuity.

(iii) The preference $\succsim$ satisfies PERU (PLRU) if and only if V is convex (concave).

(iv) The preference $\succsim$ satisfies monotonicity if and only if V is monotone (i.e., $A \subset B$ implies $V(B) \geq V(A)$ for any $A, B \in \mathcal{A}$).

Proof. Claims (iii) and (iv) follow immediately from the definitions. To prove claim (i), suppose $\succsim$ satisfies L-continuity for $M \geq 0$ and $A^*, A_* \in \mathcal{A}$. First, note that if $M = 0$, then L-continuity implies that $V(A) = V(B)$ for all $A, B \in \mathcal{A}$, i.e., V is Lipschitz continuous with a Lipschitz constant $K = 0$. If $M > 0$, then let $K \equiv 2M[V(A^*) - V(A_*)] \geq 0$. We first show that for any $A, B \in \mathcal{A}^c$,

$$d_h(A, B) \leq \frac{1}{2M} \quad \Rightarrow \quad |V(A) - V(B)| \leq K d_h(A, B). \tag{13}$$

Suppose that $d_h(A, B) \leq 1/(2M)$ and let $\alpha \equiv M d_h(A, B)$. Then $\alpha \leq 1/2$ and

$$V(B) - V(A) \leq \frac{\alpha}{1-\alpha}[V(A^*) - V(A_*)] \leq 2\alpha[V(A^*) - V(A_*)] = K d_h(A, B),$$

where the first inequality follows from L-continuity, the second inequality follows from $\alpha \leq 1/2$, and the equality follows from the definitions of α and K. Interchanging the roles of A and B above, we also have that $V(A) - V(B) \leq K d_h(A, B)$, proving (13).

Next, we use the argument in the proof of Lemma 8 in the Supplemental Material of DLRS (2007) to show that for any $A, B \in \mathcal{A}^c$,

$$|V(A) - V(B)| \leq K d_h(A, B), \tag{14}$$

i.e., the requirement $d_h(A, B) \leq 1/(2M)$ in (13) is not necessary. To see this, take any sequence $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_n < \lambda_{n+1} = 1$ such that $(\lambda_{i+1} - \lambda_i) d_h(A, B) \leq 1/(2M)$. Let $A_i = \lambda_i A + (1 - \lambda_i) B$. It is straightforward to verify that[39]

$$d_h(A_{i+1}, A_i) = (\lambda_{i+1} - \lambda_i) d_h(A, B) \leq \frac{1}{2M}.$$

Combining this with the triangle inequality and (13), we obtain

$$\begin{aligned}|V(A) - V(B)| &\leq \sum_{i=0}^{n} |V(A_{i+1}) - V(A_i)| \\ &\leq K \sum_{i=0}^{n} d_h(A_{i+1}, A_i) = K \sum_{i=0}^{n} (\lambda_{i+1} - \lambda_i) d_h(A, B) = K d_h(A, B).\end{aligned}$$

[38] If $\succsim$ also satisfies IR, then it can be shown that V is Lipschitz continuous on $\mathcal{A}$.
[39] Note that the convexity of the menus A and B is needed for the first equality.

This establishes (14), which implies V is Lipschitz continuous on $\mathcal{A}^c$ with a Lipschitz constant K.

To prove claim (ii), suppose that V is Lipschitz continuous and let $K > 0$ be a Lipschitz constant of V. Let A^* be a maximizer of V on $\mathcal{A}$ and let A_* be a minimizer of V on $\mathcal{A}$. If $V(A^*) = V(A_*)$, then $P \sim Q$ for any $P, Q \in \triangle(\mathcal{A})$, implying that L-continuity holds trivially for A^*, A_*, and $M = 0$. If $V(A^*) > V(A_*)$, then let $M \equiv K/[V(A^*) - V(A_*)] > 0$. For any $A, B \in \mathcal{A}$ and $\alpha \in [0,1]$ with $\alpha \geq M d_h(A, B)$, we have

$$(1-\alpha)[V(B) - V(A)] \leq V(B) - V(A) \leq K d_h(A, B) \leq K\alpha/M = \alpha[V(A^*) - V(A_*)],$$

which implies the conclusion of L-continuity. □

We now follow a construction similar to the one in DLR (2001) to obtain from V a function W whose domain is the set of support functions. As in the text, let

$$\mathcal{U} = \left\{ u \in \mathbb{R}^Z : \sum_{z \in Z} u_z = 0, \sum_{z \in Z} u_z^2 = 1 \right\}.$$

For any $A \in \mathcal{A}^c$, the support function $\sigma_A : \mathcal{U} \to \mathbb{R}$ of A is defined by $\sigma_A(u) = \max_{p \in A} u \cdot p$. For a more complete introduction to support functions, see Rockafellar (1970) or Schneider (1993). Let $C(\mathcal{U})$ denote the set of continuous real-valued functions on $\mathcal{U}$. When endowed with the supremum norm $\|\cdot\|_\infty$, $C(\mathcal{U})$ is a Banach space. Define an order $\geq$ on $C(\mathcal{U})$ by $f \geq g$ if $f(u) \geq g(u)$ for all $u \in \mathcal{U}$. Let $\Sigma = \{\sigma_A \in C(\mathcal{U}) : A \in \mathcal{A}^c\}$. For any $\sigma \in \Sigma$, let

$$A_\sigma = \bigcap_{u \in \mathcal{U}} \left\{ p \in \triangle(Z) : u \cdot p = \sum_{z \in Z} u_z p_z \leq \sigma(u) \right\}.$$

The next lemma summarizes some important properties of support functions. See DLR (2001) or Ergin and Saver (2010a, Lemmas 5 and 6), for precise references and additional details.

Lemma 8. *(i) For all $A \in \mathcal{A}^c$ and $\sigma \in \Sigma$, $A_{(\sigma_A)} = A$ and $\sigma_{(A_\sigma)} = \sigma$. Hence, σ is a bijection from $\mathcal{A}^c$ to Σ.*

(ii) For all $A, B \in \mathcal{A}^c$ and any $\lambda \in [0,1]$, $\sigma_{\lambda A + (1-\lambda)B} = \lambda \sigma_A + (1-\lambda)\sigma_B$.

(iii) For all $A, B \in \mathcal{A}^c$, $d_h(A, B) = \|\sigma_A - \sigma_B\|_\infty$.

(iv) We have that Σ is convex and compact, and $0 \in \Sigma$.

Using the properties of support functions established in Lemma 8, the following result shows that a function defined on $\mathcal{A}^c$ can be transformed into a function on Σ, while maintaining the properties described in Lemma 7. For a proof, see Ergin and Saver (2010a, Lemma 7).

Lemma 9. *Suppose $V : \mathcal{A}^c \to \mathbb{R}$, and define a function $W : \Sigma \to \mathbb{R}$ by $W(\sigma) = V(A_\sigma)$. Then the following statements hold:*

(i) *We have* $V(A) = W(\sigma_A)$ *for all* $A \in \mathcal{A}^c$.

(ii) *We have that* V *is Lipschitz continuous if and only if* W *is Lipschitz continuous.*

(iii) *We have that* V *is convex (concave) if and only if* W *is convex (concave).*

(iv) *We have that* V *is monotone if and only if* W *is monotone (i.e.,* $\sigma \leq \sigma'$ *implies* $W(\sigma) \leq W(\sigma')$ *for any* $\sigma, \sigma' \in \Sigma$*).*

We denote the set of continuous linear functionals on $C(\mathcal{U})$ (the dual space of $C(\mathcal{U})$) by $C(\mathcal{U})^*$. It is well known that $C(\mathcal{U})^*$ is the set of finite signed Borel measures on $\mathcal{U}$, where the duality is given by

$$\langle f, \mu \rangle = \int_{\mathcal{U}} f(u)\mu(du)$$

for any $f \in C(\mathcal{U})$ and $\mu \in C(\mathcal{U})^*$.[40]

For any function $W : \Sigma \to \mathbb{R}$, define the subdifferential ∂W and the conjugate W^* as in Appendix A. Also, define Σ_W, $\mathcal{N}_W$, and $\mathcal{M}_W$ as in (9), (10), and (11), respectively:

$$\begin{aligned} \Sigma_W &= \{\sigma \in \Sigma : \partial W(\sigma) \text{ is a singleton}\} \\ \mathcal{N}_W &= \{\mu \in C(\mathcal{U})^* : \mu \in \partial W(\sigma) \text{ for some } \sigma \in \Sigma_W\} \\ \mathcal{M}_W &= \bar{\mathcal{N}}_W, \end{aligned}$$

where the closure is taken with respect to the weak* topology. We now apply Theorem 8 to the current setting.

Lemma 10. *Suppose* $W : \Sigma \to \mathbb{R}$ *is Lipschitz continuous and convex. Then* $\mathcal{M}_W$ *is weak* compact, and for any weak* compact* $\mathcal{M} \subset C(\mathcal{U})^*$,

$$\mathcal{M}_W \subset \mathcal{M} \quad \Leftrightarrow \quad W(\sigma) = \max_{\mu \in \mathcal{M}} [\langle \sigma, \mu \rangle - W^*(\mu)] \quad \forall \sigma \in \Sigma.$$

Proof. We simply need to verify that $C(\mathcal{U})$, Σ, and W satisfy the assumptions of Theorem 8. Since $\mathcal{U}$ is a compact metric space, $C(\mathcal{U})$ is separable (see Theorem 8.48 of Aliprantis and Border 1999). By Lemma 8(iv), Σ is a closed and convex subset of $C(\mathcal{U})$ containing the origin. Since Σ is a closed subset of a Banach space, it is a Baire space by the Baire category theorem. Although the result is stated slightly differently, it is shown in Hörmander (1955) that span(Σ) is dense in $C(\mathcal{U})$. This result is also proved in DLR (2001). Since $0 \in \Sigma$ implies that aff$(\Sigma) = $ span(Σ), the affine hull of Σ is, therefore, dense in $C(\mathcal{U})$. Finally, W is Lipschitz continuous and convex by assumption. □

[40]Since $\mathcal{U}$ is a compact metric space, by the Riesz representation theorem (see Royden 1988, p. 357), each continuous linear functional on $C(\mathcal{U})$ corresponds uniquely to a finite signed Baire measure on $\mathcal{U}$. Since $\mathcal{U}$ is a locally compact separable metric space, the Baire sets and the Borel sets of $\mathcal{U}$ coincide (see Royden 1988, p. 332). Hence, the sets of Baire and Borel finite signed measures also coincide.

B.1 *Sufficiency of the axioms for the RFHA representations*

To prove the sufficiency of the axioms for the signed RFHA representation in Theorem 11.A, suppose that $\succsim$ satisfies Axiom 1, IR, and PERU. By Lemma 6, there exists a continuous function $V : \mathcal{A} \to \mathbb{R}$ such that $\mathbb{E}_P[V]$ represents $\succsim$. Moreover, by Lemma 7, the restriction of V to the set $\mathcal{A}^c$ of convex menus is Lipschitz continuous and convex. With slight abuse of notation, we also denote this restriction by V. By Lemma 9, the function $W : \Sigma \to \mathbb{R}$ defined by $W(\sigma) = V(A_\sigma)$ is Lipschitz continuous and convex. Therefore, by Lemma 10, for all $\sigma \in \Sigma$,

$$W(\sigma) = \max_{\mu \in \mathcal{M}_W} [\langle \sigma, \mu \rangle - W^*(\mu)].$$

This implies that for all $A \in \mathcal{A}$,

$$\begin{aligned} V(A) = V(\mathrm{co}(A)) &= W(\sigma_{\mathrm{co}(A)}) \\ &= \max_{\mu \in \mathcal{M}_W} \left(\int_{\mathcal{U}} \max_{p \in \mathrm{co}(A)} u(p)\mu(du) - W^*(\mu) \right) \\ &= \max_{\mu \in \mathcal{M}_W} \left(\int_{\mathcal{U}} \max_{p \in A} u(p)\mu(du) - W^*(\mu) \right), \end{aligned}$$

where the first equality follows from IR and the second equality follows from Lemma 9(i). The function W^* is lower semicontinuous by Lemma 4(i), and $\mathcal{M}_W$ is compact by Lemma 10. It is also immediate from Lemma 10 that $\mathcal{M}_W$ satisfies the minimality condition in Definition 8. Therefore, $(\mathcal{M}_W, W^*|_{\mathcal{M}_W})$ is a signed RFHA representation for $\succsim$.

To prove the sufficiency of the axioms for the (monotone) RFHA representation in Theorem 11.B, suppose that, in addition, $\succsim$ satisfies monotonicity. Then, by Lemmas 7 and 9, the function W is monotone. Also, note that for any $A, B \in \mathcal{A}^c$, $\sigma_A \vee \sigma_B = \sigma_{A \cup B}$. Hence, $\sigma \vee \sigma' \in \Sigma$ for any $\sigma, \sigma' \in \Sigma$. Therefore, by Theorem 9, the measures in $\mathcal{M}_W$ are positive.

B.2 *Sufficiency of the axioms for the MN representations*

To prove the sufficiency of the axioms for the signed MN representation in Theorem 11.A, suppose that $\succsim$ satisfies Axiom 1, IR, and PLRU. By Lemma 6, there exists a continuous function $V : \mathcal{A} \to \mathbb{R}$ such that $\mathbb{E}_P[V]$ represents $\succsim$. By Lemmas 7 and 9, the function $W : \Sigma \to \mathbb{R}$ defined by $W(\sigma) = V(A_\sigma)$ is Lipschitz continuous and concave. Define a function $\bar{W} : \Sigma \to \mathbb{R}$ by $\bar{W}(\sigma) = -W(\sigma)$. Then $\bar{W}$ is Lipschitz continuous and convex, so by Lemma 10, for all $\sigma \in \Sigma$,

$$\bar{W}(\sigma) = \max_{\mu \in \mathcal{M}_{\bar{W}}} [\langle \sigma, \mu \rangle - \bar{W}^*(\mu)].$$

Let $\mathcal{M} \equiv -\mathcal{M}_{\bar{W}} = \{-\mu : \mu \in \mathcal{M}_{\bar{W}}\}$ and define $c : \mathcal{M} \to \mathbb{R}$ by $c(\mu) = \bar{W}^*(-\mu)$. Then, for any $\sigma \in \Sigma$,

$$\begin{aligned} W(\sigma) = -\bar{W}(\sigma) &= \min_{\mu \in \mathcal{M}_{\bar{W}}} [-\langle \sigma, \mu\rangle + \bar{W}^*(\mu)] \\ &= \min_{\mu \in \mathcal{M}} [-\langle \sigma, -\mu\rangle + \bar{W}^*(-\mu)] \\ &= \min_{\mu \in \mathcal{M}} [\langle \sigma, \mu\rangle + c(\mu)]. \end{aligned}$$

This implies that for all $A \in \mathcal{A}$,

$$\begin{aligned} V(A) = V(\text{co}(A)) &= W(\sigma_{\text{co}(A)}) \\ &= \min_{\mu \in \mathcal{M}} \left(\int_{\mathcal{U}} \max_{p \in \text{co}(A)} u(p)\mu(du) + c(\mu)\right) \\ &= \min_{\mu \in \mathcal{M}} \left(\int_{\mathcal{U}} \max_{p \in A} u(p)\mu(du) + c(\mu)\right), \end{aligned}$$

where the first equality follows from IR and the second equality follows from Lemma 9(i). The function $\bar{W}^*$ is lower semicontinuous by Lemma 4(i), which implies that c is lower semicontinuous. The compactness of $\mathcal{M}$ follows from the compactness of $\mathcal{M}_{\bar{W}}$, which follows from Lemma 10. Also, by Lemma 10 and the above construction, it is immediate that $\mathcal{M}$ satisfies the minimality condition in Definition 8. Therefore, $(\mathcal{M}, c)$ is a signed MN representation for $\succsim$.

To prove the sufficiency of the axioms for the (monotone) MN representation in Theorem 11.B, suppose that, in addition, $\succsim$ satisfies monotonicity. Then, by Lemmas 7 and 9, the function W is monotone. Let $\hat{\Sigma} \equiv -\Sigma = \{-\sigma : \sigma \in \Sigma\}$ and define a function $\hat{W} : \hat{\Sigma} \to \mathbb{R}$ by $\hat{W}(\sigma) \equiv \bar{W}(-\sigma) = -W(-\sigma)$. Notice that $\hat{W}$ is monotone and convex: By the monotonicity of W, for any $\sigma, \sigma' \in \hat{\Sigma}$,

$$\sigma \leq \sigma' \quad \Rightarrow \quad -\sigma \geq -\sigma' \quad \Rightarrow \quad \hat{W}(\sigma) = -W(-\sigma) \leq -W(-\sigma') = \hat{W}(\sigma).$$

By the concavity of W, for any $\sigma, \sigma' \in \hat{\Sigma}$ and $\lambda \in [0, 1]$,

$$\begin{aligned} \hat{W}(\lambda\sigma + (1-\lambda)\sigma') &= -W(\lambda(-\sigma) + (1-\lambda)(-\sigma')) \\ &\leq -\lambda W(-\sigma) - (1-\lambda)W(-\sigma') = \lambda\hat{W}(\sigma) + (1-\lambda)\hat{W}(\sigma'). \end{aligned}$$

Also, for any $A, B \in \mathcal{A}^c$, $(-\sigma_A) \wedge (-\sigma_B) = -(\sigma_A \vee \sigma_B) = -\sigma_{A\cup B}$. Hence, $\sigma \wedge \sigma' \in \hat{\Sigma}$ for any $\sigma, \sigma' \in \hat{\Sigma}$. Therefore, by Theorem 9, the measures in $\mathcal{M}_{\hat{W}}$ are positive. For any $\mu \in C(\mathcal{U})^*$ and $\sigma, \sigma' \in \hat{\Sigma}$, note that

$$\hat{W}(\sigma') - \hat{W}(\sigma) \geq \langle \sigma' - \sigma, \mu\rangle \quad \Leftrightarrow \quad \bar{W}(-\sigma') - \bar{W}(-\sigma) \geq \langle \sigma' - \sigma, \mu\rangle = \langle -\sigma' + \sigma, -\mu\rangle$$

and, hence, $\mu \in \partial\hat{W}(\sigma) \Leftrightarrow -\mu \in \partial\bar{W}(-\sigma)$. In particular, $\hat{\Sigma}_{\hat{W}} = -\Sigma_{\bar{W}}$ and $\mathcal{N}_{\hat{W}} = -\mathcal{N}_{\bar{W}}$. Taking closures, we have $\mathcal{M}_{\hat{W}} = -\mathcal{M}_{\bar{W}} = \mathcal{M}$. Thus, the measures in $\mathcal{M}$ are positive.

B.3 *Necessity of the axioms*

We begin by demonstrating some of the properties of the function V defined by a signed representation.

Lemma 11. *Suppose $(\mathcal{M}, c)$ is a signed representation.*

(i) If $(\mathcal{M}, c)$ is a signed RFHA representation and $V : \mathcal{A} \to \mathbb{R}$ is defined by (4), then V is Lipschitz continuous and convex. In addition, defining the function $W : \Sigma \to \mathbb{R}$ by $W(\sigma) = V(A_\sigma)$, we have $\mathcal{M} = \mathcal{M}_W$ and $c(\mu) = W^*(\mu)$ for all $\mu \in \mathcal{M}$.

(ii) If $(\mathcal{M}, c)$ is a signed MN representation and $V : \mathcal{A} \to \mathbb{R}$ is defined by (12), then V is Lipschitz continuous and concave. In addition, defining the function $W : \Sigma \to \mathbb{R}$ by $W(\sigma) = V(A_\sigma)$, we have $\mathcal{M} = -\mathcal{M}_{-W}$ and $c(\mu) = [-W]^*(-\mu)$ for all $\mu \in \mathcal{M}$.

Proof. (i) By the definitions of V and W, we have

$$W(\sigma) = \max_{\mu \in \mathcal{M}} [\langle \sigma, \mu \rangle - c(\mu)] \quad \forall \sigma \in \Sigma.$$

By Theorem 10(i), W is Lipschitz continuous and convex. Therefore, the restriction of V to $\mathcal{A}^c$ is Lipschitz continuous and convex by Lemma 9. Let $K \geq 0$ be any Lipschitz constant of $V|_{\mathcal{A}^c}$, and take any $A, B \in \mathcal{A}$. It is easily verified that $V(A) = V(\mathrm{co}(A))$, $V(B) = V(\mathrm{co}(B))$ and $d_h(\mathrm{co}(A), \mathrm{co}(B)) \leq d_h(A, B)$. Hence,

$$|V(A) - V(B)| = \big|V(\mathrm{co}(A)) - V(\mathrm{co}(B))\big| \leq K d_h(\mathrm{co}(A), \mathrm{co}(B)) \leq K d_h(A, B),$$

which implies that V is Lipschitz continuous on all of $\mathcal{A}$ with the same Lipschitz constant K. Also, for any $A, B \in \mathcal{A}$ and $\lambda \in [0, 1]$,

$$\begin{aligned} V(\lambda A + (1-\lambda)B) &= V(\mathrm{co}(\lambda A + (1-\lambda)B)) = V(\lambda\, \mathrm{co}(A) + (1-\lambda)\, \mathrm{co}(B)) \\ &\leq \lambda V(\mathrm{co}(A)) + (1-\lambda)V(\mathrm{co}(B)) = \lambda V(A) + (1-\lambda)V(B), \end{aligned}$$

which implies that V is convex on $\mathcal{A}$. Also, by Theorem 10(ii) and (iii) and the compactness of Σ, $\mathcal{M}_W \subset \mathcal{M}$ and $W^*(\mu) = c(\mu)$ for all $\mu \in \mathcal{M}_W$. By Lemma 10 and the minimality of $\mathcal{M}$, this implies $\mathcal{M} = \mathcal{M}_W$, and, hence, $c(\mu) = W^*(\mu)$ for all $\mu \in \mathcal{M}$.

(ii) Define a function $\bar{W} : \Sigma \to \mathbb{R}$ by $\bar{W}(\sigma) = -W(\sigma)$. Then, for any $\sigma \in \Sigma$,

$$\begin{aligned} \bar{W}(\sigma) = -W(\sigma) &= -\min_{\mu \in \mathcal{M}} [\langle \sigma, \mu \rangle + c(\mu)] \\ &= \max_{\mu \in \mathcal{M}} [\langle \sigma, -\mu \rangle - c(\mu)] \\ &= \max_{\mu \in -\mathcal{M}} [\langle \sigma, \mu \rangle - c(-\mu)]. \end{aligned}$$

By the same arguments used above, this implies that $\bar{W}$ is Lipschitz continuous and convex, which in turn implies that V is Lipschitz continuous and concave. Moreover, the above arguments imply that $-\mathcal{M} = \mathcal{M}_{\bar{W}}$ and $c(-\mu) = \bar{W}^*(\mu)$ for all $\mu \in -\mathcal{M}$. Thus, $\mathcal{M} = -\mathcal{M}_{\bar{W}} = -\mathcal{M}_{-W}$ and $c(\mu) = \bar{W}^*(-\mu) = [-W]^*(-\mu)$ for all $\mu \in \mathcal{M}$. $\square$

Suppose that $\succsim$ has a signed RFHA (signed MN) representation $(\mathcal{M}, c)$, and suppose $V : \mathcal{A} \to \mathbb{R}$ is defined by (4) ((12)). Since $\mathbb{E}_P[V]$ represents $\succsim$ and V is continuous (by Lemma 11), $\succsim$ satisfies weak order, continuity, and first-stage independence by Lemma 6. Since V is Lipschitz continuous and convex (concave) by Lemma 11, $\succsim$ satisfies L-continuity and PERU (PLRU) by Lemma 7. Since $V(A) = V(\text{co}(A))$ for all $A \in \mathcal{A}$, it is immediate that $\succsim$ satisfies IR. Finally, if the measures in $\mathcal{M}$ are positive, then it is obvious that V is monotone, which implies that $\succsim$ satisfies monotonicity.

Appendix C: Proof of Theorem 1

Assume first that there exists an RFHA representation $(\mathcal{M}, c)$ such that V is given by (4). By Lemma 1(ii), there exists $((\Omega, \mathcal{F}, \pi), \Theta, U)$ as in an HA representation where $\Theta = \mathcal{M}$, such that for all $\mu \in \mathcal{M}$ and $A \in \mathcal{A}$,

$$\int_{\mathcal{U}} \max_{p \in A} u(p)\mu(du) = \int_{\Omega} \max_{p \in A} U(p, \omega; \mu)\pi(d\omega; \mu).$$

It follows that the V satisfies (3) for the HA representation $((\Omega, \mathcal{F}, \pi), \Theta, U, c)$ with the cost function c.

For the converse, assume that there exists an HA representation such that V is given by (3). It is easy to see that V is monotone, convex, and satisfies $V(A) = V(\text{co}(A))$ for all $A \in \mathcal{A}$. An argument similar to the one in Appendix D.1 in Ergin and Saver (2010a) can be used to also show that boundedness of U implies that V is Lipschitz continuous. Therefore, the construction in Appendix B.1 implies that there exists an RFHA representation such that V is given by (4).

Appendix D: Proof of Theorem 3

We next state and prove a generalization of Theorem 3 to signed representations (see Definitions 8 and 9). Theorem 3 is a special case of Theorem 12 and, therefore, follows directly.

Theorem 12. *If $(\mathcal{M}, c)$ and $(\mathcal{M}', c')$ are two signed RFHA (signed MN) representations for $\succsim$, then there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $\mathcal{M}' = \alpha\mathcal{M}$ and $c'(\alpha\mu) = \alpha c(\mu) + \beta$ for all $\mu \in \mathcal{M}$.*

Proof. Throughout the proof, we will continue to use notation and results for support functions that were established in Appendix B. Suppose $(\mathcal{M}, c)$ and $(\mathcal{M}', c')$ are two signed RFHA representations for $\succsim$. Define $V : \mathcal{A} \to \mathbb{R}$ and $V' : \mathcal{A} \to \mathbb{R}$ for these respective representations, and define $W : \Sigma \to \mathbb{R}$ and $W' : \Sigma \to \mathbb{R}$ by $W(\sigma) = V(A_\sigma)$ and $W'(\sigma) = V'(A_\sigma)$. By Lemma 11(i), $\mathcal{M} = \mathcal{M}_W$ and $c(\mu) = W^*(\mu)$ for all $\mu \in \mathcal{M}$. Similarly, $\mathcal{M}' = \mathcal{M}_{W'}$ and $c'(\mu) = W'^*(\mu)$ for all $\mu \in \mathcal{M}'$.

Since V is continuous (by Lemma 11), the uniqueness part of Lemma 6 implies that there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $V' = \alpha V - \beta$. This implies that $W' = \alpha W - \beta$. Therefore, for any $\sigma, \sigma' \in \Sigma$,

$$W(\sigma') - W(\sigma) \geq \langle \sigma' - \sigma, \mu \rangle \quad \Leftrightarrow \quad W'(\sigma') - W'(\sigma) \geq \langle \sigma' - \sigma, \alpha\mu \rangle$$

and, hence, $\partial W'(\sigma) = \alpha\partial W(\sigma)$. In particular, $\Sigma_{W'} = \Sigma_W$ and $\mathcal{N}_{W'} = \alpha\mathcal{N}_W$. Taking closures, we also have that $\mathcal{M}_{W'} = \alpha\mathcal{M}_W$. Since from our earlier arguments $\mathcal{M}' = \mathcal{M}_{W'}$ and $\mathcal{M} = \mathcal{M}_W$, we conclude that $\mathcal{M}' = \alpha\mathcal{M}$. Finally, let $\mu \in \mathcal{M}$. Then

$$c'(\alpha\mu) = \sup_{\sigma\in\Sigma}[\langle\sigma, \alpha\mu\rangle - W'(\sigma)] = \alpha \sup_{\sigma\in\Sigma}[\langle\sigma, \mu\rangle - W(\sigma)] + \beta = \alpha c(\mu) + \beta,$$

where the first and last equalities follow from our earlier findings that $c' = W'^*|_{\mathcal{M}_{W'}}$ and $c = W^*|_{\mathcal{M}_W}$.

The proof of the uniqueness of the signed MN representation is similar and involves an application of Lemma 11(ii). $\square$

Appendix E: Proof of Theorem 5

We define the set of *translations* to be

$$\Theta \equiv \left\{\theta \in \mathbb{R}^Z : \sum_{z\in Z} \theta_z = 0\right\}.$$

For $A \in \mathcal{A}$ and $\theta \in \Theta$, define $A + \theta \equiv \{p + \theta : p \in A\}$. Intuitively, adding θ to A in this sense simply "shifts" A. Also note that for any $p, q \in \triangle(Z)$, we have $p - q \in \Theta$.

Definition 10. A function $V : \mathcal{A} \to \mathbb{R}$ is *translation linear* if there exists $v \in \mathbb{R}^Z$ such that for all $A \in \mathcal{A}$ and $\theta \in \Theta$ with $A + \theta \in \mathcal{A}$, we have $V(A + \theta) = V(A) + v \cdot \theta$.

Lemma 12. *Suppose that $V : \mathcal{A} \to \mathbb{R}$ is a function such that $\mathbb{E}_P[V]$ represents the preference $\succsim$ over $\triangle(\mathcal{A})$. Then V is translation linear if and only if $\succsim$ satisfies RDD.*

Proof. Assume that $\mathbb{E}_P[V]$ represents the preference $\succsim$. Then it is easy to see that $\succsim$ satisfies RDD if and only if

$$V(\alpha A + (1-\alpha)\{p\}) - V(\alpha A + (1-\alpha)\{q\}) = (1-\alpha)\big[V(\{p\}) - V(\{q\})\big] \tag{15}$$

for any $\alpha \in [0, 1]$, $A \in \mathcal{A}$, and $p, q \in \triangle(Z)$.

If there exists $v \in \mathbb{R}^Z$ such that for all $A \in \mathcal{A}$ and $\theta \in \Theta$ with $A + \theta \in \mathcal{A}$, we have $V(A + \theta) = V(A) + v \cdot \theta$, then both sides of (15) are equal to $(1-\alpha)v \cdot (p - q)$, showing that $\succsim$ satisfies RDD.

If $\succsim$ satisfies RDD, then define the function $f : \triangle(Z) \to \mathbb{R}$ by $f(p) = V(\{p\})$ for all $p \in \triangle(Z)$. Let $\alpha \in [0, 1]$ and $p, q \in \triangle(Z)$. Then

$$\begin{aligned}
2f(\alpha p + (1-\alpha)q) &= \big[f(\alpha p + (1-\alpha)q) - f(\alpha p + (1-\alpha)p)\big] \\
&\qquad + \big[f(\alpha p + (1-\alpha)q) - f(\alpha q + (1-\alpha)q)\big] + f(p) + f(q) \\
&= (1-\alpha)[f(q) - f(p)] + \alpha[f(p) - f(q)] + f(p) + f(q) \\
&= 2[\alpha f(p) + (1-\alpha)f(q)],
\end{aligned}$$

where the second equality follows from (15) and the definition of f. Therefore, $f(\alpha p+(1-\alpha)q)=\alpha f(p)+(1-\alpha)f(q)$ for any $\alpha\in[0,1]$ and $p,q\in\triangle(Z)$. It is standard to show that this implies that there exists $v\in\mathbb{R}^Z$ such that $f(p)=v\cdot p$ for all $p\in\triangle(Z)$.

To see that V is translation linear, let $A\in\mathcal{A}$ and $\theta\in\Theta$ be such that $A+\theta\in\mathcal{A}$. If $\theta=0$, then the conclusion of translation linearity follows trivially, so without loss of generality assume that $\theta\neq 0$. Ergin and Saver (2010a) show in the proof of their Lemma 4 that if $A\in\mathcal{A}$ and $A+\theta\in\mathcal{A}$ for some $\theta\in\Theta\setminus\{0\}$, then there exist $A'\in\mathcal{A}$, $p,q\in\triangle(Z)$, and $\alpha\in(0,1]$ such that $A=(1-\alpha)A'+\alpha\{p\}$, $A+\theta=(1-\alpha)A'+\alpha\{q\}$, and $\theta=\alpha(p-q)$. Then

$$\begin{aligned}V(A+\theta)-V(A)&=V((1-\alpha)A'+\alpha\{p\})-V((1-\alpha)A'+\alpha\{q\})\\&=\alpha\big[V(\{p\})-V(\{q\})\big]\\&=\alpha[v\cdot p-v\cdot q]\\&=v\cdot\theta,\end{aligned}$$

where the second equality follows from (15) and the third equality follows from the expected utility form of f. Therefore, V is translation linear. □

We are now ready to prove Theorem 5. The necessity of RDD is straightforward and left to the reader. For the other direction, suppose that $\succsim$ has an RFHA representation $(\mathcal{M},c)$ and that it satisfies RDD. In the rest of this section, we will continue to use the notation and results from Appendix B. By Theorem 2, $\succsim$ satisfies Axiom 1 and PERU. Therefore, $(\mathcal{M}_W, W^*|_{\mathcal{M}_W})$ constructed in Appendix B is also an RFHA representation for $\succsim$. Since $\succsim$ satisfies RDD, by Lemma 12, the value function V for this representation is translation linear. Let $v\in\mathbb{R}^Z$ be such that for all $A\in\mathcal{A}$ and $\theta\in\Theta$ with $A+\theta\in\mathcal{A}$, we have $V(A+\theta)=V(A)+v\cdot\theta$. Let $q=(1/|Z|,\dots,1/|Z|)\in\triangle(Z)$. By Lemma 22 of Ergin and Saver (2010a), for all $\mu\in\mathcal{M}_W$ and $p\in\triangle(Z)$, $\langle\sigma_{\{p\}},\mu\rangle=v\cdot(p-q)$. The consistency of $\mathcal{M}_W$ follows immediately from this fact because for any $\mu,\mu'\in\mathcal{M}_W$ and $p\in\triangle(Z)$, we have

$$\int_{\mathcal{U}}u(p)\mu(du)=\langle\sigma_{\{p\}},\mu\rangle=v\cdot(p-q)=\langle\sigma_{\{p\}},\mu'\rangle=\int_{\mathcal{U}}u(p)\mu'(du).$$

By Theorem 3, there exists $\alpha>0$ such that $\mathcal{M}=\alpha\mathcal{M}_W$. Therefore, $(\mathcal{M},c)$ is also consistent.

Appendix F: Proof of Theorem 6

F.1 *Proof of Theorem 6.A*

The necessity of the axioms is straightforward. For sufficiency, suppose that $\succsim$ satisfies Axiom 1, mixture independence, and monotonicity. By Lemma 6, there exists a continuous function $V:\mathcal{A}\to\mathbb{R}$ such that $P\succsim Q$ if and only if $\mathbb{E}_P[V]\geq\mathbb{E}_Q[V]$.

Define a preference $\succsim'$ on $\mathcal{A}^c$ by $A\succsim' B\Leftrightarrow\delta_A\succsim\delta_B$ (or, equivalently, $A\succsim' B\Leftrightarrow V(A)\geq V(B)$). The axioms assumed on $\succsim$ then imply that $\succsim'$ satisfies the DLR (2001) axioms: Continuity of $\succsim$ implies continuity of $\succsim'$; mixture independence implies that $\succsim'$

satisfies independence; and monotonicity of $\succsim$ implies that $\succsim'$ satisfies monotonicity. Therefore, by Theorem S2 in the Supplemental Material of DLRS (2007), there exists a finite Borel measure μ on $\mathcal{U}$ such that $U:\mathcal{A}\to\mathbb{R}$ defined by

$$U(A)=\int_{\mathcal{U}}\max_{p\in A}u(p)\mu(du)$$

represents $\succsim'$. Moreover, since U is continuous and $\mathcal{A}$ is compact, there exist $-\infty < a \leq b < +\infty$ such that $[a,b]=\{U(A):A\in\mathcal{A}\}$. Since $V(A)\geq V(B)\Leftrightarrow U(A)\geq U(B)$, there exists a strictly increasing function $\phi:[a,b]\to\mathbb{R}$ such that

$$V(A)=\phi(U(A)).$$

To establish the Lipschitz continuity of ϕ, first recall that by Lemma 7, L-continuity implies there exists $K\geq 0$ such that $|V(A)-V(B)|\leq Kd_h(A,B)$ for any $A,B\in\mathcal{A}^c$ (the set of all convex menus). If $a=b$, then ϕ is trivially Lipschitz continuous. Next suppose that $a<b$. Take $A_*, A^*\in\mathcal{A}^c$ such that $U(A_*)=a$ and $U(A^*)=b$. For any $t\in[a,b]$, let $\alpha(t)\equiv(t-a)/(b-a)\in[0,1]$, which implies $U(\alpha(t)A^*+(1-\alpha(t))A_*)=t$. Note that for any $\alpha,\beta\in[0,1]$,

$$d_h(\alpha A^*+(1-\alpha)A_*,\beta A^*+(1-\beta)A_*)=|\alpha-\beta|d_h(A^*,A_*).$$

Thus, for any $s,t\in[a,b]$,

$$\begin{aligned}|\phi(t)-\phi(s)| &= \left|\phi\big(U\big(\alpha(t)A^*+(1-\alpha(t))A_*\big)\big)-\phi\big(U\big(\alpha(s)A^*+(1-\alpha(s))A_*\big)\big)\right|\\ &= \left|V\big(\alpha(t)A^*+(1-\alpha(t))A_*\big)-V\big(\alpha(s)A^*+(1-\alpha(s))A_*\big)\right|\\ &\leq K|\alpha(t)-\alpha(s)|d_h(A^*,A_*)\\ &= K|t-s|d_h(A^*,A_*)/(b-a),\end{aligned}$$

which implies ϕ is Lipschitz continuous with a Lipschitz constant of $Kd_h(A^*,A_*)/(b-a)$.

F.2 *Proof of Theorem 6.B*

The necessity of the axioms is straightforward. For sufficiency, suppose that $\succsim$ satisfies Axiom 1, second-stage independence, and strategic rationality. By Theorem 6.A, $\succsim$ has a KPDLR representation (ϕ,μ). It therefore suffices to show that μ has singleton support.

Suppose, to the contrary, that μ has more than one utility function in its support. Fix any $u',u''\in\operatorname{supp}(\mu)$. Choose lotteries $p,q\in\triangle(Z)$ such that $u'(p)>u'(q)$ and $u''(q)>u''(p)$. Then since these inequalities also hold on any small neighborhoods of u' and u'', respectively, this implies

$$\int_{\mathcal{U}}\max_{r\in\{p,q\}}u(r)\mu(du)>\int_{\mathcal{U}}u(p)\mu(du)\quad\text{and}\quad\int_{\mathcal{U}}\max_{r\in\{p,q\}}u(r)\mu(du)>\int_{\mathcal{U}}u(q)\mu(du).$$

Therefore, $V(\{p,q\})>V(\{p\})$ and $V(\{p,q\})>V(\{q\})$, which constitutes a violation of the strategic rationality axiom. Thus, if $\succsim$ satisfies strategic rationality, $\operatorname{supp}(\mu)=\{u\}$ for some $u\in\mathcal{U}$. Taking $\alpha=\mu(\{u\})$, we then have $\mu=\alpha\delta_u$, as desired.

F.3 *Proof of Theorem 6.C*

Suppose $\succsim$ has a KPDLR representation (ϕ, μ). First, note that for any $A, B \in \mathcal{A}$ and any $\alpha \in (0,1)$,

$$\int_{\mathcal{U}} \max_{p \in \alpha A + (1-\alpha)B} u(p)\mu(du) = \alpha \int_{\mathcal{U}} \max_{p \in A} u(p)\mu(du) + (1-\alpha)\int_{\mathcal{U}} \max_{p \in B} u(p)\mu(du).$$

For any $s, t \in [a,b]$, let $A, B \in \mathcal{A}$ be menus that satisfy $s = \int_{\mathcal{U}} \max_{p\in A} u(p)\mu(du)$ and $t = \int_{\mathcal{U}} \max_{p \in B} u(p)\mu(du)$. Then, for any $\alpha \in (0,1)$,

$$\begin{aligned}
&\alpha\delta_A + (1-\alpha)\delta_B \succsim \delta_{\alpha A + (1-\alpha)B} \\
&\quad\Leftrightarrow \quad \alpha V(A) + (1-\alpha)V(B) \geq V(\alpha A + (1-\alpha)B) \\
&\quad\Leftrightarrow \quad \alpha\phi\left(\int_{\mathcal{U}} \max_{p\in A} u(p)\mu(du)\right) + (1-\alpha)\phi\left(\int_{\mathcal{U}} \max_{p\in B} u(p)\mu(du)\right) \\
&\qquad\qquad\qquad \geq \phi\left(\alpha \int_{\mathcal{U}} \max_{p\in A} u(p)\mu(du) + (1-\alpha)\int_{\mathcal{U}} \max_{p \in B} u(p)\mu(du)\right) \\
&\quad\Leftrightarrow \quad \alpha\phi(s) + (1-\alpha)\phi(t) \geq \phi(\alpha s + (1-\alpha)t).
\end{aligned}$$

Thus, $\succsim$ satisfies PERU if and only if ϕ is convex. A similar argument shows that $\succsim$ satisfies PLRU if and only if ϕ is concave.

Appendix G: Proof of Theorem 7

Throughout this section, we use the notation ∂f, f^*, $\mathcal{N}_f$, and $\mathcal{M}_f$ introduced in Appendix A.

Lemma 13. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $\phi : [a,b] \to \mathbb{R}$ be Lipschitz continuous and convex. Then $1 \Leftrightarrow 2 \Rightarrow 3$:*

(i) We have that ϕ is strictly increasing.

(ii) (a) We have $\mathcal{M}_\phi \subset \mathbb{R}_+$.

(b) The right derivative of ϕ^* at 0, $(d\phi^*/d\lambda)(0+)$, exists and is equal to a.

(iii) We have that 0 is not an isolated point of $\mathcal{M}_\phi$.

Proof. (i) $\Rightarrow$ (ii) Part (ii)(a) follows from Theorem 9.

To see part (ii)(b), it is enough to show that for all $t \in (a,b]$, there exists $\lambda > 0$ such that

$$\lambda' a \leq \phi^*(\lambda') - \phi^*(0) \leq \lambda' t \quad \forall \lambda' \in (0, \lambda). \tag{16}$$

Since ϕ is nondecreasing, $0 \in \partial\phi(a)$. Along with Lemma 4, this implies that $-\phi^*(0) = \phi(a) \geq \lambda' a - \phi^*(\lambda')$ for any $\lambda' \geq 0$, establishing the first inequality in (16). Take any

$t \in (a, b]$. By Lemma 3 there exists $\lambda \in \partial\phi(t)$. Note that $\lambda > 0$. Otherwise, if $\lambda \leq 0$, then by Lemma 4,

$$\phi(a) \geq \lambda a - \phi^*(\lambda) \geq \lambda t - \phi^*(\lambda) = \phi(t),$$

a contradiction to ϕ being strictly increasing. Let $\lambda' \in (0, \lambda)$. Since ϕ is continuous and its domain is compact, there exists $t' \in [a, b]$ such that $\phi^*(\lambda') = t'\lambda' - \phi(t')$. By Lemma 4, this implies that $\lambda' \in \partial\phi(t')$. Monotonicity of the subdifferential $\partial\phi$ implies that $t' \leq t$.[41] Then, by Lemma 4 and ϕ being nondecreasing,

$$-\phi^*(0) = \phi(a) \leq \phi(t') = \lambda' t' - \phi^*(\lambda') \leq \lambda' t - \phi^*(\lambda'),$$

which implies the second inequality in (16).

(ii) $\Rightarrow$ (i) Theorem 8 and part (ii)(a) imply that ϕ is nondecreasing. Therefore, $0 \in \partial\phi(a)$, implying $\phi(a) = -\phi^*(0)$ by Lemma 4.

We will first show that $\phi(a) < \phi(t)$ for any $t \in (a, b]$. Suppose for a contradiction that $\phi(a) = \phi(t)$ for some $t \in (a, b]$. Then, for any $\lambda > 0$,

$$\phi^*(\lambda) \geq \lambda t - \phi(t) = \lambda t - \phi(a) = \lambda t + \phi^*(0),$$

implying that $(\phi^*(\lambda) - \phi^*(0))/\lambda \geq t > a$ for any $\lambda > 0$, a contradiction to $(d\phi^*/d\lambda)(0+) = a$.

To conclude that ϕ is strictly increasing, it remains to show that $\phi(t) < \phi(t')$ for any $t, t' \in (a, b]$ such that $t < t'$. By Lemma 3, there exists $\lambda \in \partial\phi(t)$. If $\lambda \leq 0$, then

$$\phi(a) \geq \lambda a - \phi^*(\lambda) \geq \lambda t - \phi^*(\lambda) = \phi(t)$$

by Lemma 4, contradicting $\phi(a) < \phi(t)$. Therefore, $\lambda > 0$, implying

$$\phi(t) = \lambda t - \phi^*(\lambda) < \lambda t' - \phi^*(\lambda) \leq \phi(t'),$$

by Lemma 4, as desired.

(i) $\Rightarrow$ (iii) Suppose, to the contrary, that 0 is an isolated point of $\mathcal{M}_\phi$. Then $0 \in \mathcal{N}_\phi$, i.e., there exists $t \in [a, b]$ such that $\partial\phi(t) = \{0\}$. Then Lemma 4 implies

$$-\phi^*(0) = \phi(t) > \lambda t - \phi^*(\lambda) \quad \forall \lambda \in \mathcal{M}_\phi \setminus \{0\}.$$

Since 0 is an isolated point of $\mathcal{M}_\phi$ and $\mathcal{M}_\phi$ is compact by Theorem 8, $\mathcal{M}_\phi \setminus \{0\}$ is also compact. Therefore, the above inequality implies that

$$-\phi^*(0) > \max_{\lambda \in \mathcal{M}_\phi \setminus \{0\}} [\lambda t - \phi^*(\lambda)]. \tag{17}$$

Let $\Delta > 0$ be the difference of the left-hand side and the right-hand side in (17), and let $M > 0$ be such that $\mathcal{M}_\phi \subset [0, M]$. Take any $s \in [a, b]$ such that $|t - s| < \Delta/M$. Then

[41] To see that $\partial\phi$ is monotone, note that by the definition of the subdifferential, $\lambda \in \partial\phi(t)$ implies $\lambda(t' - t) \leq \phi(t') - \phi(t)$ and $\lambda' \in \partial\phi(t')$ implies $\lambda'(t - t') \leq \phi(t) - \phi(t')$. Summing these inequalities, we have $(\lambda - \lambda')(t - t') \geq 0$.

$|\lambda t - \lambda s| < \Delta$ for any $\lambda \in \mathcal{M}_\phi \setminus \{0\}$, implying that (17) continues to hold if t is replaced by s. Therefore,

$$-\phi^*(0) = \max_{\lambda \in \mathcal{M}_\phi} [\lambda s - \phi^*(\lambda)] = \phi(s),$$

where the second equality follows from Theorem 8. This implies that ϕ is constant at a Δ/M neighborhood of t, contradicting the assumption that ϕ is strictly increasing. $\square$

In the next lemma, Σ denotes the set of support functions defined in Appendix B.

Lemma 14. *Let μ be a nonzero finite signed Borel measure on $\mathcal{U}$ and let $[a, b] = \{\langle\sigma, \mu\rangle : \sigma \in \Sigma\}$. Let $\phi : [a, b] \to \mathbb{R}$ be Lipschitz continuous and convex, and define $W : \Sigma \to \mathbb{R}$ by $W(\sigma) = \phi(\langle\sigma, \mu\rangle)$ for any $\sigma \in \Sigma$. Then the following statements hold:*

(i) We have that W is Lipschitz continuous and convex.

(ii) We have $W^*(\lambda\mu) = \phi^*(\lambda)$ for any $\lambda \in \mathbb{R}$.

(iii) We have $\mathcal{M}_W = \{\lambda\mu : \lambda \in \mathcal{M}_\phi\}$.

Proof. (i) Let $K \geq 0$ be a Lipschitz constant for ϕ. Then, for any $\sigma, \sigma' \in \Sigma$,

$$|W(\sigma) - W(\sigma')| = \big|\phi(\langle\sigma, \mu\rangle) - \phi(\langle\sigma', \mu\rangle)\big| \leq K|\langle\sigma, \mu\rangle - \langle\sigma', \mu\rangle| \leq K\|\mu\|\|\sigma - \sigma'\|,$$

implying that W is Lipschitz continuous with a Lipschitz constant $K\|\mu\|$. The variable W is convex as the composition of a linear and a convex function.

(ii) Let $\lambda \in \mathbb{R}$. Then

$$\begin{aligned} W^*(\lambda\mu) &= \max_{\sigma \in \Sigma} [\langle\sigma, \lambda\mu\rangle - W(\sigma)] \\ &= \max_{\sigma \in \Sigma} \big[\lambda\langle\sigma, \mu\rangle - \phi(\langle\sigma, \mu\rangle)\big] \\ &= \max_{t \in [a,b]} [\lambda t - \phi(t)] \\ &= \phi^*(\lambda). \end{aligned}$$

(iii) We will first show that $\mathcal{N}_W \subset \{\lambda\mu : \lambda \in \mathcal{M}_\phi\}$. This will imply that $\mathcal{M}_W = \bar{\mathcal{N}}_W \subset \{\lambda\mu : \lambda \in \mathcal{M}_\phi\}$ since $\mathcal{M}_\phi$ is closed. Let $\nu \in \mathcal{N}_W$. Then there exists $\sigma \in \Sigma$ such that $\partial W(\sigma) = \{\nu\}$. For any $\lambda \in \partial\phi(\langle\sigma, \mu\rangle)$,

$$W(\sigma') - W(\sigma) = \phi(\langle\sigma', \mu\rangle) - \phi(\langle\sigma, \mu\rangle) \geq \lambda[\langle\sigma', \mu\rangle - \langle\sigma, \mu\rangle] = \langle\sigma' - \sigma, \lambda\mu\rangle \quad \forall \sigma' \in \Sigma,$$

implying $\lambda\mu \in \partial W(\sigma) = \{\nu\}$. Therefore, $\{\lambda\mu : \lambda \in \partial\phi(\langle\sigma, \mu\rangle)\} \subset \{\nu\}$. Since μ is nonzero and $\partial\phi(\langle\sigma, \mu\rangle) \neq \varnothing$ by Lemma 3, there exists a unique $\lambda \in \mathbb{R}$ such that $\partial\phi(\langle\sigma, \mu\rangle) = \{\lambda\}$. Note that $\lambda \in \mathcal{N}_\phi \subset \mathcal{M}_\phi$ and $\nu = \lambda\mu$, as desired.

Let $\mathcal{M} = \{\lambda \in \mathbb{R} : \lambda\mu \in \mathcal{M}_W\}$. We will next show that $\mathcal{M}_\phi \subset \mathcal{M}$, which will imply $\{\lambda\mu : \lambda \in \mathcal{M}_\phi\} \subset \mathcal{M}_W$. Since μ is nonzero and $\mathcal{M}_W$ is compact by part (i) and Theorem 8, $\mathcal{M}$ is also compact. Let $t \in [a, b]$ and let $\sigma \in \Sigma$ be such that $t = \langle\sigma, \mu\rangle$. Then

$$\phi(t) = W(\sigma) = \max_{\nu \in \mathcal{M}_W} [\langle\sigma, \nu\rangle - W^*(\nu)] = \max_{\lambda \in \mathcal{M}} [\langle\sigma, \lambda\mu\rangle - W^*(\lambda\mu)] = \max_{\lambda \in \mathcal{M}} [\lambda t - \phi^*(\lambda)],$$

where the second equality follows from part (i) and Theorem 8, the third equality follows from $\mathcal{M}_W \subset \{\lambda\mu : \lambda \in \mathbb{R}\}$, and the last equality follows from part (ii). Therefore, by Theorem 8, $\mathcal{M}_\phi \subset \mathcal{M}$. $\square$

Proof of Theorem 7. In the following discussion, let $W : \Sigma \to \mathbb{R}$ be defined by $W(\sigma) = V(A_\sigma)$. Also, let $[a, b] = \{\int_{\mathcal{U}} \max_{p\in A} u(p)\mu(du) : A \in \mathcal{A}\}$.

(i) $\Rightarrow$ (ii) For any $\sigma \in \Sigma$,

$$W(\sigma) = V(A_\sigma) = \phi(\langle \sigma_{(A_\sigma)}, \mu\rangle) = \phi(\langle \sigma, \mu\rangle),$$

where the last equality follows from Lemma 8(i). Since W is Lipschitz continuous and convex by Lemma 14, $V(A) = V(\text{co}(A))$ for all $A \in \mathcal{A}$, and $W(\sigma) = V(A_\sigma)$ for all $\sigma \in \Sigma$, the construction in Appendix B.1 implies that $(\mathcal{M}, c) := (\mathcal{M}_W, W^*|_{\mathcal{M}_W})$ is an RFHA representation such that V is given by (4). By Lemma 13(ii)(a) and Lemma 14(iii), $\mathcal{M}_W \subset \{\lambda\mu : \lambda \in \mathbb{R}_+\}$. By Lemma 13(ii)(b) and Lemma 14(ii),

$$\lim_{\lambda \searrow 0} \frac{W^*(\lambda\mu) - W^*(0)}{\lambda} = \frac{d\phi^*}{d\lambda^+}(0) = a \equiv \min_{A\in\mathcal{A}} \int_{\mathcal{U}} \max_{p\in A} u(p)\mu(du). \tag{18}$$

By Lemma 13(iii), Lemma 14(iii), and μ being nonzero, 0 is not an isolated point of $\mathcal{M}_W$. Therefore, if $0 \in \mathcal{M}_W$, then the limit term in (18) agrees with $\lim_{\lambda\searrow 0 :\, \lambda\mu\in\mathcal{M}_W} (c(\lambda\mu) - c(0))/\lambda$.

(ii) $\Rightarrow$ (i) The mapping $\lambda \mapsto c(\lambda\mu)$ is lower semicontinuous since c is lower semicontinuous, and $\{\lambda \in \mathbb{R}_+ : \lambda\mu \in \mathcal{M}\}$ is nonempty by Theorem 7(ii)(a), and it is compact since $\mathcal{M}$ is compact and μ is nonzero. Therefore, we can define $\phi : [a, b] \to \mathbb{R}$ by

$$\phi(t) = \max_{\lambda\in\mathbb{R}_+ :\, \lambda\mu\in\mathcal{M}} [\lambda t - c(\lambda\mu)] \quad \forall t \in [a, b].$$

By Theorem 10, ϕ is Lipschitz continuous and convex. Furthermore, for any $A \in \mathcal{A}$,

$$V(A) = \max_{\lambda\in\mathbb{R}_+ :\, \lambda\mu\in\mathcal{M}} [\langle \sigma_A, \lambda\mu\rangle - c(\lambda\mu)] = \max_{\lambda\in\mathbb{R}_+ :\, \lambda\mu\in\mathcal{M}} [\lambda\langle \sigma_A, \mu\rangle - c(\lambda\mu)] = \phi(\langle \sigma_A, \mu\rangle),$$

where the first equality follows from (4) and Theorem 7(ii)(a). Therefore, it only remains to show that ϕ is strictly increasing.

By Lemma 11, $\mathcal{M} = \mathcal{M}_W$ and $c(\nu) = W^*(\nu)$ for all $\nu \in \mathcal{M}$. Note that

$$W(\sigma) = V(A_\sigma) = \phi(\langle \sigma_{(A_\sigma)}, \mu\rangle) = \phi(\langle \sigma, \mu\rangle) \quad \forall \sigma \in \Sigma,$$

where the last equality follows from Lemma 8(i). By Lemma 14(iii), $\mathcal{M} = \{\lambda\mu : \lambda \in \mathcal{M}_\phi\}$. Therefore, since μ is nonzero: $0 \in \mathcal{M}$ if and only if $0 \in \mathcal{M}_\phi$; Theorem 7(ii)(a) implies $\mathcal{M}_\phi \subset \mathbb{R}_+$; and the first part of Theorem 7(ii)(b) implies that 0 is not an isolated point of $\mathcal{M}_\phi$.

First suppose that $0 \notin \mathcal{M}$, implying $0 \notin \mathcal{M}_\phi$. Let $t, t' \in [a, b]$ be such that $t < t'$. By Theorem 8,

$$\phi(s) = \max_{\lambda\in\mathcal{M}_\phi} [\lambda s - \phi^*(\lambda)] \quad \forall s \in [a, b].$$

Let $\hat{\lambda} > 0$ be a solution of the above maximization at $s = t$. Then

$$\phi(t) = \hat{\lambda}t - \phi^*(\hat{\lambda}) < \hat{\lambda}t' - \phi^*(\hat{\lambda}) \leq \max_{\lambda \in \mathcal{M}_\phi} [\lambda t' - \phi^*(\lambda)] = \phi(t').$$

Next suppose that $0 \in \mathcal{M}$, implying that $0 \in \mathcal{M}_\phi$. Then

$$a = \lim_{\lambda \searrow 0 : \lambda\mu \in \mathcal{M}} \frac{c(\lambda\mu) - c(0)}{\lambda} = \lim_{\lambda \searrow 0 : \lambda \in \mathcal{M}_\phi} \frac{\phi^*(\lambda) - \phi^*(0)}{\lambda}, \tag{19}$$

where the first equality follows from Theorem 7(ii)(b) and the second equality follows from $\mathcal{M} = \{\lambda\mu : \lambda \in \mathcal{M}_\phi\}$, μ being nonzero, $c = W^*|_{\mathcal{M}}$, and Lemma 14(ii). For any $\lambda \in (0, \infty)$, define $a_\lambda \in \mathbb{R}$ by $a_\lambda = (\phi^*(\lambda) - \phi^*(0))/\lambda$. Since $0 \in \mathcal{M}_\phi$ is not an isolated point of $\mathcal{M}_\phi$, (19) implies that there exists a sequence λ_n in $\mathcal{M}_\phi \setminus \{0\}$ such that $\lambda_n \searrow 0$ and $\lim_n a_{\lambda_n} = a$. Since ϕ^* is convex, a_λ is nondecreasing in $\lambda \in (0, \infty)$. Therefore, for any sequence λ'_n in $(0, \infty)$ such that $\lambda'_n \searrow 0$, $\lim_n a_{\lambda'_n} = \lim_n a_{\lambda_n}$. This implies that the limit on the right-hand side of (19) is equal to $(d\phi^*/d\lambda)(0+)$. By Lemma 13, ϕ is strictly increasing. $\square$

References

Ahn, David S. and Todd Sarver (2013), "Preference for flexibility and random choice." *Econometrica*, 81, 341–361. [514, 516, 517]

Aliprantis, Charalambos D. and Kim C. Border (1999), *Infinite Dimensional Analysis*. Springer, Berlin. [520, 526]

Athey, Susan and Ilya Segal (2013), "An efficient dynamic mechanism." *Econometrica*, 81, 2463–2485. [496]

Backus, David K., Bryan R. Routledge, and Stanley E. Zin (2004), "Exotic preferences for macroeconomics." In *NBER Macroeconomics Annual*, Vol. 19 (Mark Gertler and Keneth Rogoff, eds.), 319–390, MIT Press, Cambridge, Massachusetts. [490]

Caplin, Andrew and John Leahy (2001), "Psychological expected utility theory and anticipatory feelings." *Quarterly Journal of Economics*, 116, 55–79. [501]

Clark, Stephen A. (1996), "The random utility model with an infinite choice space." *Economic Theory*, 7, 179–189. [516]

Compte, Olivier and Philippe Jehiel (2007), "Auctions and information acquisition: Sealed bid or dynamic formats?" *Rand Journal of Economics*, 38, 355–372. [496]

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini (2001), "Representing preferences with a unique subjective state space." *Econometrica*, 69, 891–934. [489, 491, 492, 493, 501, 505, 510, 511, 514, 516, 517, 525, 526, 532]

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini (2009), "Temptation driven preferences." *Review of Economic Studies*, 76, 937–971. [501, 521]

Dekel, Eddie, Barton L. Lipman, Aldo Rustichini, and Todd Sarver (2007), "Representing preferences with a unique subjective state space: A corrigendum." *Econometrica*, 75, 591–600. [498, 524, 533]

Ekeland, Ivar and Thomas Turnbull (1983), *Infinite-Dimensional Optimization and Convexity*. University of Chicago Press, Chicago. [519]

Epstein, Larry G. (2008), "Living with risk." *Review of Economic Studies*, 75, 1121–1141. [501]

Epstein, Larry G., Massimo Marinacci, and Kyoungwon Seo (2007), "Coarse contingencies and ambiguity." *Theoretical Economics*, 2, 355–394. [497, 505, 522]

Epstein, Larry G. and Kyoungwon Seo (2009), "Subjective states: A more robust model." *Games and Economic Behavior*, 67, 408–427. [497]

Epstein, Larry G. and Stanley E. Zin (1989), "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework." *Econometrica*, 57, 937–969. [490]

Epstein, Larry G. and Stanley E. Zin (1991), "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis." *Journal of Political Economy*, 99, 263–286. [490]

Ergin, Haluk (2003), "Costly contemplation." Unpublished paper, MIT. [507]

Ergin, Haluk and Todd Saver (2010a), "A unique costly contemplation representation." *Econometrica*, 78, 1285–1339. [492, 493, 495, 498, 502, 503, 506, 507, 519, 520, 523, 525, 530, 532]

Ergin, Haluk and Todd Sarver (2010b), "The unique minimal dual representation of a convex function." *Journal of Mathematical Analysis and Applications*, 370, 600–606. [519, 520]

Falmagne, Jean-Claude (1978), "A representation theorem for finite random scale systems." *Journal of Mathematical Psychology*, 18, 52–72. [516]

Fishburn, Peter C. (1970), *Utility Theory for Decision Making*. Wiley, New York. [523]

Gale, David (1967), "A geometric duality theorem with economic applications." *Review of Economic Studies*, 34, 19–24. [518]

Grant, Simon, Atsushi Kajii, and Ben Polak (1998), "Intrinsic preference for information." *Journal of Economic Theory*, 83, 233–259. [499]

Grant, Simon, Atsushi Kajii, and Ben Polak (2000), "Temporal resolution of uncertainty and recursive non-expected utility models." *Econometrica*, 68, 425–434. [499]

Gul, Faruk and Wolfgang Pesendorfer (2001), "Temptation and self-control." *Econometrica*, 69, 1403–1435. [501, 521]

Gul, Faruk and Wolfgang Pesendorfer (2006), "Random expected utility." *Econometrica*, 74, 121–146. [514, 516, 517]

Holmes, Richard B. (1975), *Geometric Functional Analysis and Its Applications*. Springer, New York. [519]

Hörmander, Lars (1955), "Sur la fonction d'appui des ensembles convexes dans un espace localement convexe." *Arkiv För Matematik*, 3, 181–186. [526]

Kraus, Alan and Jacob S. Sagi (2006), "Inter-temporal preference for flexibility and risky choice." *Journal of Mathematical Economics*, 42, 698–709. [510]

Kreps, David M. (1979), "A representation theorem for 'preference for flexibility'." *Econometrica*, 47, 565–577. [489, 491, 501, 512]

Kreps, David M. (1988), *Notes on the Theory of Choice*. Westview Press, Boulder, Colorado. [492, 502, 523]

Kreps, David M. and Evan Porteus (1978), "Temporal resolution of uncertainty and dynamic choice theory." *Econometrica*, 46, 185–200. [489, 490, 491, 492, 493, 495, 497, 498, 499, 500, 510, 512, 513]

Kreps, David M. and Evan L. Porteus (1979), "Temporal von Neumann–Morgenstern and induced preferences." *Journal of Economic Theory*, 20, 81–109. [490, 491, 493, 513]

Luce, R. Duncan (1959), *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York. [516]

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006), "Ambiguity aversion, robustness, and the variational representation of preferences." *Econometrica*, 74, 1447–1498. [522]

Machina, Mark J. (1984), "Temporal risk and the nature of induced preferences." *Journal of Economic Theory*, 33, 199–231. [491, 493, 494]

Markowitz, Harry M. (1959), *Portfolio Selection: Efficient Diversification of Investments*. Wiley, New York. [513]

McFadden, Daniel and Marcel K. Richter (1990), "Stochastic rationality and revealed stochastic preference." In *Preferences, Uncertainty and Optimality: Essays in Honor of Leonid Hurwicz* (John S. Chipman, Daniel McFadden, and Marcel K. Richter, eds.), 161–186, Westview Press, Boulder, Colorado. [516]

Mossin, Jan (1969), "A note on uncertainty and preferences in a temporal context." *American Economic Review*, 59, 172–174. [513]

Munkres, James R. (2000), *Topology*, second edition. Pearson, New York. [523]

Ortoleva, Pietro (2013), "The price of flexibility: Towards a theory of thinking aversion." *Journal of Economic Theory*, 148, 903–934. [507]

Phelps, Robert R. (1993), *Convex Functions, Monotone Operators, and Differentiability*, second edition. Springer, Berlin. [506]

Rockafellar, R. Tyrrell (1970), *Convex Analysis*. Princeton University Press, Princeton, New Jersey. [506, 519, 525]

Royden, Halsey L. (1988), *Real Analysis*, third edition. Macmillan, New York. [526]

Sarver, Todd (2008), "Anticipating regret: Why fewer options may be better." *Econometrica*, 76, 263–305. [498, 521]

Sarver, Todd (2014), "Optimal reference points and anticipation." Unpublished paper, Duke University. [501]

Schneider, Rolf (1993), *Convex Bodies: The Brunn–Minkowski Theory*. Cambridge University Press, Cambridge. [523, 525]

Spence, Michael and Richard Zeckhauser (1972), "The effect of the timing of consumption decisions and the resolution of lotteries on the choice of lotteries." *Econometrica*, 40, 401–403. [513]

Tallarini, Thomas D. (2000), "Risk-sensitive real business cycles." *Journal of Monetary Economics*, 45, 507–532. [490]

Weil, Philippe (1993), "Precautionary savings and the permanent income hypothesis." *Review of Economic Studies*, 60, 367–383. [490]

Submitted 2012-9-14. Final version accepted 2014-5-16. Available online 2014-5-16.