Anticipated Stochastic Choice

Nobuo Koida*

This version: November 5, 2010

Abstract

This paper characterizes the decision maker whose choice over menus, i.e., sets of alternatives, is deterministic while choice of an alternative from a menu is *stochastic*. A key idea of this paper is to consider *perfectly correlated mixtures* of menus, i.e., the mixture of two menus includes only the mixtures of particular pairs of alternatives. Our main theorem axiomatizes an *anticipated stochastic choice (ASC)* representation, in which a *state-generating function*, or a probability measure over *mental states*, is chosen from a closed and convex set so that the expected utility is maximized. ASC representations accommodate choice anomalies such as the attraction effect, cyclical choice, and Allais paradox. Special cases of ASC representations are trembling hands, i.e., the best alternative is chosen from a menu with probability close to 1 while arbitrary suboptimal alternatives are chosen with positive probability, and choice under limited consideration sets, i.e., the DM maximizes her utility by only considering alternatives in the unions of supports of state-generating functions. Moreover, the sets of state-generating functions in ASC representations are unique, thus they can be interpreted as indices of a preference for commitment to singleton menus.

Keywords: preferences over menus, stochastic choice, trembling hands

JEL Classification Numbers: D81

*Faculty of Policy Studies, Iwate Prefectural University, 152-52 Aza-Sugo, Takizawa, Takizawa-mura, Iwate 020-0193, Japan. E-mail: nobuo@iwate-pu.ac.jp, TEL: +81-19-694-2814, FAX: +81-19-694-2701

1 Introduction

1.1 Outline and Background

Does a decision maker (DM) prefer a larger or smaller *menu*, i.e., feasible set of alternatives? Not to mention traditional consumer theory, we usually assume that larger menus are (at least weakly) preferred since they provide more *flexibility* ex post; larger menus enable the DM to depend her choice of alternatives on unforeseen contingencies that are resolved ex post. This intuition is formalized by Koopmans (1964), and axiomatized by authors such as Kreps (1979, 1992), Nehring (1999), and Dekel, Lipman, and Rustichini (2001).

The objective of this paper is to demonstrate that this intuition does not necessarily hold if the DM *stochastically* chooses an alternative from a menu, rather than deterministically: suppose that the DM is asked to choose a restaurant that serves both chicken and fish or that serves only chicken. She has a strict preference for chicken over fish, but when arriving at the former restaurant, she may end up choosing the unwanted fish dish for some reason – its description is typed in red so that it attracts her attention, she is in a rush and chooses it without intention, or she forgets her original plan when she chooses an alternative.

Since we can naturally regard such choice behavior as stochastic, ex post choice is also regarded as stochastic. Then, fully anticipating this ex ante, she may prefer *commitment* to a singleton menu. Namely, she prefers the restaurant that serves only chicken to the one that serves both chicken and fish, to avoid "mistakenly" choosing suboptimal alternatives ex post. This paper formalizes this intuition.

A key assumption of the characterization is that mixtures of menus are *perfectly correlated*, i.e., they include only mixtures of particular pairs of alternatives. This is the case if the common set of underlying *mental states* determine the choice of alternatives from each menu. This clearly contrasts with uncorrelated mixtures of menus employed in the existing studies, which include mixtures of all pairs of alternatives from each menu.

Our main theorem (Theorem 1) axiomatizes the following preference representation, which we refer to as an *anticipated stochastic choice (ASC) representation*:

$$V(x) = \max_{\mu \in \mathcal{M}_{|x|}} \int_{S_{|x|}} u(\phi(x,s))d\mu(s), \tag{1}$$

where x is a menu, $S_{|x|}$ is the set of mental states that depends on the cardinality of x, $\mathcal{M}_{|x|}$ is a set of probability measures over $S_{|x|}$, or *state-generating functions*, $\mathcal{S} = \cup_{n=1}^{\infty}\mathcal{S}_n$, $\mathcal{M} = \cup_{n=1}^{\infty}\mathcal{M}_n$, u is a (state-independent) utility function, and ϕ is a *choice function*, i.e., a function that yields an alternative of a given menu, depending on a mental state s_i. Probability distributions over the alternatives of a menu derived from a choice function ϕ and state-generating functions μ can be interpreted as stochastic (probabilistic) choice functions.

An important property of an ASC representation is the uniqueness of $\mathcal{M}$. Hence, $\mathcal{M}$ is naturally interpreted as an index of bounded rationality, or ability to control her future self. If $\mathcal{M}_{|x|} = \Delta(S_{|x|})$, i.e., the set of state-generating functions is equivalent to the set of all probability distributions over $S_{|x|}$, then she has full control of her future choice. In this case, the DM can always attain the exact maximization, i.e., $V(x) = \max_{\beta \in x} u(\beta)$, which is roughly corresponds to a preference for flexibility (with a state-independent utility function). On the other hand, if $\mathcal{M}_{|x|} = \{\mu\}$ for some probability measure μ over $S_{|x|}$, then her ex post choice is purely stochastic. This case corresponds to Luce's (1959) choice axiom approach. ASC representations include these two as polar cases.

Moreover, there are two more notable special cases of an ASC representation. One is a *trembling-hand* model (Selten 1975), which implies that the DM chooses the best alternative with probability $1-\epsilon$ and suboptimal alternatives with probability ϵ. The other is choice under limited *consideration sets* (Wright and Barbour 1977; Masatlioglu, Nakajima, and Ozbay 2010), i.e., the DM exactly maximizes her utility except that only alternatives within a specific set are considered.

To summarize, the contributions of this paper are threefold. First, we demonstrate that stochastic choice is a (potential) source of a preference for commitment, in addition to temptation (Gul and Pesendorfer 2001; Dekel, Lipman, and Rustichini 2009; Dekel and Lipman 2010; Stovall 2010; Chatterjee and Krishna 2009; Noor and Takeoka 2009, 2010) and ex post regret (Sarver 2008).

Second, we endogenize a wide range of stochastic choice models from a preference for menus. Although many studies have discussed stochastic choice, they directly assume, rather than derive from a preference, a probability measure over alternatives or related objects (e.g., Debreu 1958; Luce 1959; Davidson and Marschak 1959; Luce and Suppes 1965; Selten 1975; McFadden and Richter 1991; Harless and Camerer 1994; Bandyopadhyay et al. 1999; Gul and Pesendorfer 2006; Manzini and Mariotti 2010). In contrast, we *derive* a subjective probability measure over alternatives from a preference for menus.

Finally, an ASC representation is more flexible than stochastic choice models described by a single probability measure over alternatives, so that it does not necessarily satisfy regularity (Luce and Suppes 1965) or Weak Axiom of Revealed Stochastic Preference (WARSP, Bandyopadhyay et al. 1999). Hence, it accommodates anomalies of stochastic choice such as the attraction effect, cyclical choice, and Allais paradox. (We will discuss this in Section 4.)

The remainder of this paper is organized as follows. In the next subsection, we review the related literature. In Section 2, we provide the basic framework that we focus on. In Section 3, we provide basic axioms, and in Section 4, we propose the main representation theorem (Theorem 1). Section 5 characterizes the relation between preferences over menus of different cardinalities. We examine the implication of ASC representations regarding size-related axioms (such as monotonicity) in Section 6 and the pure stochastic case, where the set of state-generating function is a singleton, in Section 7. We discuss ASC representations accommodating some anomalies of stochastic or deterministic choice and the implications of mental states compared to the literature in Section 8. Concluding remarks are made in Section 9.

1.2 Related Literature

In this section, we will review the relevant literature. First, our model is related to menu choice models, in particular, the models of a preference for flexibility and commitment (Kreps 1979, 1992; Nehring 1999; Dekel, Lipman, and Rustichini 2001, 2009; Gul and Pesendorfer 2001; Sarver 2008; Chatterjee and Krishna 2009; Noor and Takeoka 2009, 2010; Stovall 2010; Ergin and Sarver 2010; Dekel and Lipman 2010). Three remarks are in order.

The first is that an ASC representation does *not* generally satisfy *extremeness* (Gul and Pesendorfer 2006), i.e., an ASC representation generally permits non-extreme points of (the convex hull of) a menu to be chosen with positive probability. On the other hand, as Gul and Pesendorfer point out, the studies mentioned above *do* satisfy extremeness since they are random utility models. Accordingly, our ASC representation outperforms these studies as a model of stochastic choice, because extremeness is violated in many experiments.

The second (somewhat technical but related to the first) remark is that we employ perfectly correlated mixtures of menus whereas they employ uncorrelated mixtures. The difference leads to representations over choice functions in this paper and over support functions in the studies mentioned above, which changes the relevance of extremeness.

The final remark is on Ergin and Sarver's (2010, Section 5.1) constant (in particular, zero) contemplation cost representation. The reader may suspect that their representation looks similar to an ASC representation in that the optimal probability measure over "subjective" states is chosen from a set, as well as their set of axioms are similar to ours. However, our model differs from theirs for the following two reasons. First, as we mentioned, their model satisfies extremeness, which is difficult to interpret as a stochastic choice model. Second, Ergin and Sarver's representation is compatible with monotonicity, i.e., menu x is preferred to y if $x \supseteq y$, whereas our representation contradicts it unless it reduces to the exact maximization case as we will indicate in Section 6.

Next, we discuss the relations to various stochastic choice models. First, ASC representations are generally more flexible than a stochastic choice model described by a single probability measure over alternatives (e.g., Luce 1959) since they admit more than one state-generating functions. This property enables them to accommodate choice anomalies as we will demonstrate in Section 8. Second, this paper provides a decision theoretic foundation for trembling-hand models, which are employed in game theory and experimental economics (e.g., Selten 1975; Harless and Camerer 1994). Moreover, Theorem 4 in Section 6 also provides a foundation for the argument that the less rational the DM becomes (parameter ϵ increases), the higher the probability of choosing suboptimal alternatives becomes.

Choice under limited consideration sets was originally considered in psychology and marketing science and has also been studied in economics (for extensive references, see Masatlioglu, Nakajima, and Ozbay 2010). As we mentioned in the previous section, if the union of supports of state-generating functions is a proper subset of the set of all mental states, the corresponding ASC representation can be interpreted as (exact) utility maximization under limited consideration sets, whereby only specific (e.g., N most salient) alternatives are chosen from a menu with positive probability. On the other hand, Manzini and Mariotti (2010) assume stochastic consideration sets and derive a stochastic choice model from them. Unlike their approach, we first derive a stochastic choice model from a menu preference, then derive consideration sets in a special case of ASC representations.

The separation between ex ante optimality and ex post choice/optimality assumed in this paper resembles the dichotomy of *deliberative* and *affective* (and often automatic) systems (Loewenstein and O'donoghue 2004). Similar dichotomies prevail in economics (Thaler and Shefrin 1981; Loewenstein 1996; Kahneman 1994; Kahneman, Wakker, and Sarin 1997; Gul and Pesendorfer 2001; Fudenberg and Levine 2006; Chatterjee and Krishna 2009), as well as psychology and neuroscience. From this point of view, an ASC representation describes a situation in which the affective system triggers stochastic ex post choice, whereas fully anticipating this, the DM deliberatively chooses a menu and controls the realization of mental states to maximize her ex ante (expected) utility.

2 Preliminaries

Let Z be a finite set of prizes and $\Delta(Z)$ the set of lotteries, i.e., Borel measures over Z, or alternatives. The symbol $\mathcal{K}(\cdot)$ denotes the set of all subsets of $\Delta(Z)$, where $\mathcal{K}_0(\cdot)$ denotes the set of all finite subsets. We set $\mathcal{A} = \mathcal{K}_0(\Delta(Z))$, the set of finite subsets of $\Delta(Z)$, as the domain of choice in this paper.[1] We denoted by $\mathcal{A}_M$ the set of menus with cardinality $M = 1, 2, \cdots$.

[1] We restrict our attention to finite menus since a key assumption of our analysis is that there exists a bijection (or onto) function between alternatives in each menu, which is not generally satisfied with (uncountably) infinite menus. Note that our approach can be easily extended to the case with countable menus. On the other hand, many studies focus on finite menus or convex hulls of finite menus (Gul and Pesendorfer 2001; Dekel, Lipman,

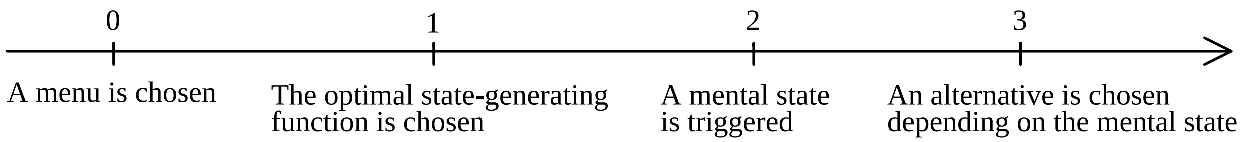

Figure 1: Timeline

We apply the topology to $\mathcal{A}$ that is induced by the Hausdorff (semi-)metric

$$d_h(x, y) = \max\{\max_{\beta \in x} \min_{\gamma \in y} d_p(\beta, \gamma), \max_{\gamma \in y} \min_{\beta \in x} d_p(\beta, \gamma)\},$$

where x, $y \in \mathcal{A}$ and d_p is a metric that generates the weak topology.[2]

In this paper, we assume a preference relation $\succsim$ on $\mathcal{A}$. We refer to the restriction of $\succsim$ to singleton menus, i.e., a binary relation on $\Delta(Z)$, as the *commitment ranking*. For simplicity, we denote $\beta \succsim \gamma$ to indicate $\{\beta\} \succsim \{\gamma\}$ for β, $\gamma \in \Delta(Z)$.

2.1 Timeline

Now, we explain the timeline assumed in this paper.[3] At period 0, the DM chooses a menu x. Then, at period 1, the DM chooses a state-generating function μ, i.e., a probability measure over mental states, anticipating her own ex post stochastic choice, to maximize the ex ante expected utility. At period 2, a (subjective) mental state s, which will govern choice at period 3, is triggered. Finally, the DM chooses an alternative β from menu x at period 3, depending on the mental state s.

Unlike the existing studies on unforeseen contingencies (e.g., Dekel, Lipman, and Rustichini 2001), we do *not* necessarily assume a utility maximizing stage after the realization of a mental state (this is the reason that we apply the terminology "mental state" rather than "subjective state," which is often employed in the literature): even if β_2 is not preferred to β_1 by the commitment ranking, it can be chosen from menu $\{\beta_1, \beta_2\}$ with positive probability.

There are two possible interpretations for this assumption. The first views that ex post choice as an (automatic) process that is irrelevant to utility maximization, despite the DM is fully aware of this ex ante.[4] This interpretation reflects the dichotomy between the deliberative

and Rustichini 2009; Stovall 2010; Chatterjee and Krishna 2009).

[2]Since we focus on finite menus in this paper, both the maxima and minima exist.

[3]The timeline detailed here is an interpretation, rather than a part of the model, since we assume that only choice over menus at period 1 is observable.

[4]Some psychology literature reported that the subjects' estimations of the probability to make a specific form of mistakes conform with their partners', are steady in the long run, and correlate with those of the probability to make other forms of mistakes, despite they still cannot avoid making the mistakes in practice (e.g., Broadbent et al. 1982).

and affective systems in decision making that we discussed in Subsection 1.2, where the ex ante preference corresponds to the former while the ex post preference corresponds to the latter.

In contrast, the second interpretation regards that each mental state renders the corresponding alternative optimal in terms of the ex post preference, i.e., each alternative is the (unique) maximizer of the ex post utility given some mental state s. According to this view, ex ante suboptimal alternatives are chosen ex post because of the discrepancy between the ex ante and ex post preferences, rather than non-utility-maximizing behavior such as mistakes.[5]

In the remainder of this paper, we follow the first interpretation unless otherwise stated, since our main focus is stochastic choice rather than random utility, for which the first interpretation is more relevant.

2.2 Correlated Mixtures

In this paper, we employ *correlated* mixtures of menus, instead of uncorrelated mixtures employed by the existing studies. We first provide an intuition. Suppose that each alternative is associated with unique salience that is irrelevant to the prizes that an alternative induces, e.g., the description of each alternative is printed in letters with different colors;[6] the DM is vulnerable to specific salience in some mental state – she chooses an alternative printed in red in the mood that it attracts her attention, regardless of the prizes that it gives. Then, we can reasonably assume that the DM chooses alternatives with the same type of salience in the common mental state, i.e., she chooses an alternative printed in red from a menu whenever she does from another menu. Since the timeline implies that mental states are realized before the randomization between menus x and y of mixture $\lambda x+(1-\lambda)y$ is performed, the DM naturally identifies the mixture with a menu that only includes mixtures of correlated alternatives. Conversely, if she always identifies the mixture of menus with a menu that only includes mixtures of correlated alternatives, we can construct the set of mental states that trigger the choice of the corresponding alternatives from a menu.

In the following, we formalize this idea. First, we coin $\rho_{x,y} : x \times y \to \{0,1\}$ as the *mixture coefficient for menus x and y* and define $\lambda x+(1-\lambda)y \equiv \{\lambda\beta+(1-\lambda)\gamma : \beta \in x, \gamma \in y, \rho_{x,y}(\beta,\gamma) = 1\}$. In other words, $\rho_{x,y}(\beta,\gamma) = 1$ indicates that the mixture of alternatives β and γ is included in the mixture of menus x and y, while $\rho_{x,y}(\beta,\gamma) = 0$ indicates not.[7] We say that alternatives

[5]This could also lead to a multi-agent interpretation, where ex ante and ex post preferences describe a "principal" and an "agent," respectively, rather than multiple selves of a single DM.

[6]It is possible that salience depends on the prizes that an alternative induces. For instance, many experiments suggest that extreme (maximum and/or minimum) outcomes be crucial in some choice problems such as the Allais paradox, in which case salience is no longer independent of the outcomes that an alternative induces. We exclude these cases for simplicity.

[7]Note that $\rho_{x,y}(\beta,\gamma)$ depends on members of x and y other than β and γ. Namely, it is possible that $\rho_{x,y}(\beta,\gamma) = 1$ and $\rho_{x',y'}(\beta,\gamma) = 0$ for some menus x, y, x', and y' such that $\beta \in x$, $\beta \in x'$, $\gamma \in y$, and $\gamma \in y'$.

β and γ in menus x and y, respectively, are *mixed* if $\rho_{x,y}(\beta, \gamma) = 1$. Note that uncorrelated mixtures in the literature correspond to the case that $\rho_{x,y}(\beta, \gamma) = 1$ for all x, $y \in \mathcal{A}$, $\beta \in x$, and $\gamma \in y$.

We impose the following conditions on mixture coefficients throughout this paper.

Definition 1 (Perfectly Correlated Mixture Coefficient) A mixture coefficient ρ is called a *perfectly correlated mixture coefficient* if the following conditions hold:
(a) $\rho_{x,y}(\beta, \gamma) = \rho_{y,x}(\gamma, \beta)$ for all x, $y \in \mathcal{A}$, $\beta \in x$ and $\gamma \in y$;
(b) $\rho_{x,y}(\beta, \gamma) = 1$ and $\rho_{y,z}(\gamma, \delta) = 1$ imply $\rho_{x,z}(\beta, \delta) = 1$ for all x, y, $z \in \mathcal{A}$, $|x| = |y| = |z|$ and $\beta \in x$, $\gamma \in y$, and $\delta \in z$;
(c) a function $\xi_{x,y} : \Delta(Z) \to \Delta(Z)$ such that $\xi_{x,y}(\beta) = \gamma$ whenever $\rho_{x,y}(\beta, \gamma) = 1$ for all $\beta \in x$, $\gamma \in y$ is onto for all x, $y \in \mathcal{A}$ with $|x| \geq |y|$;
(d) $\rho_{x,y}(\beta, \gamma) = 1$ implies $\rho_{x,\lambda x+(1-\lambda)y}(\beta, \lambda\beta + (1 - \lambda)\gamma) = 1$ for all x, $y \in \mathcal{A}$, $\beta \in x$, $\gamma \in y$, and $\lambda \in [0, 1]$.

Condition (a) implies symmetry, i.e., alternative β in menu x is mixed with γ in y whenever γ in y is mixed with β in x. Condition (b) implies transitivity, i.e., if alternative β in menu x is mixed with γ in menu y and δ in menu z is mixed with δ in menu z, then β in x and δ in z are mixed, with menus with the same cardinality. Condition (c) implies that for all x, y with $|x| \geq |y|$, each alternative in the menu x is mixed with exactly one alternative in the menu y while each alternative in y is mixed with at least one alternative in x. Note that conditions (a) to (c) render the mixture of menus well-defined. On the other hand, Condition (d) implies that the correlation between alternatives is invariant with respect to the convex combinations of alternatives, i.e., if β and γ are correlated, then β (and γ by condition (a)) is mixed with the mixture of β and γ.

To illustrate the difference between perfectly correlated and uncorrelated mixtures, consider menus $x = \{\beta_1, \beta_2\}$ and $y = \{\gamma_1, \gamma_2\}$. If mixtures are uncorrelated as in the existing studies, i.e., $\rho_{x,y}(\beta, \gamma) = 1$ for all $\beta \in x$ and $\gamma \in y$, the mixture $\lambda x + (1 - \lambda)y$ with $\lambda \in [0, 1]$ is equivalent to $\{\lambda\beta_1 + (1 - \lambda)\gamma_1, \lambda\beta_1 + (1 - \lambda)\gamma_2, \lambda\beta_2 + (1 - \lambda)\gamma_1, \lambda\beta_2 + (1 - \lambda)\gamma_2\}$. On the other hand, if the mixture is perfectly correlated and $\rho_{x,y}(\beta_i, \gamma_i) = 1$ for $i = 1$, 2, it is equivalent to $\{\lambda\beta_1 + (1 - \lambda)\gamma_1, \lambda\beta_2 + (1 - \lambda)\gamma_2\}$. Note that a perfectly correlated mixture is equivalent to an uncorrelated mixture if one (or both) of the two menus is a singleton.

Next, we define a useful property of an alternative, which we call a mixture-invariant addi-

This reflects the fact that salience is often relative rather than absolute: the "redness" of a color is relatively determined by other colors that the DM faces (orange is more red than yellow, but less red than red), which may change the the mental state that triggers the choice of the alternative in the color.

tion.

Definition 2 (Mixture-Invariant Addition) For $x \in \mathcal{A}$, we call $\tilde{\beta} \in \Delta(Z)$ a *mixture-invariant addition to* x if there exists $y \in \mathcal{A}$ and $\tilde{\gamma} \in \Delta(Z)$ such that $\rho_{x\cup\{\tilde{\beta}\},y\cup\{\tilde{\gamma}\}}(\beta,\gamma) = 1$ whenever $\rho_{x,y}(\beta,\gamma) = 1$, and $\rho_{x\cup\{\tilde{\beta}\},y\cup\{\tilde{\gamma}\}}(\tilde{\beta},\tilde{\gamma}) = 1$. Further, given such y and $\tilde{\gamma}$, we call $\tilde{\beta}$ a *mixture-invariant addition to* x *with respect to* $(y,\tilde{\gamma})$.

A mixture-invariant addition to a menu remains the mixtures of existing alternatives unchanged, while it is mixed to the alternative that is added to the other menu. Note that the alternative $\tilde{\gamma}$ in Definition 2 is a mixture-invariant addition to y by definition.

To obtain the main result, we need to make the following structural assumption.

Assumption 1 (a) for all $x \in \mathcal{A}$, $\bar{\beta} \in x$, and $\beta' \in \Delta(Z)$, there exists $\hat{\beta} \in \Delta(Z)$ such that $\hat{\beta} \sim \beta'$ and $\hat{\beta}$ is a mixture-invariant addition to $x \setminus \{\bar{\beta}\}$ with respect to $(x \setminus \{\bar{\beta}\}, \bar{\beta})$, and
(b) for all $x \in \mathcal{A}$, there exists a mixture-invariant addition to x.

Condition (a) assures that an arbitrary alternative can be replaced by the alternative of an arbitrary commitment ranking without changing the mixtures between alternatives, whereas (b) states that there exists an alternative such that its addition to a menu does not change the mixture between alternatives.

The following lemma, whose proof is relegated in the Appendix, shows that a perfectly correlated mixture coefficient derives the sets of mental states and a choice function.

Lemma 1 *Under Assumption 1b, the following statements are equivalent:*
(a) ρ *is a perfectly correlated mixture coefficient.*
(b) for all M*, there exist* $S_M = \{s_1, \cdots, s_M\}$ *and* $\phi : \mathcal{A}_M \times S_M \to \Delta(Z)$ *such that* $\phi(x,s) \in x$ *for all* $s \in S_M$ *and* $\phi(x,s) \neq \phi(x,s')$ *for* $s \neq s'$*. Further, functions* $\phi_x(\cdot) \equiv \phi(x,\cdot)$ *for some* $x \in \mathcal{A}_M$ *are closed under state-wise mixtures, i.e.,* $(\lambda\phi_x + (1-\lambda)\phi_y)(s) \equiv \lambda\phi_x(s) + (1-\lambda)\phi_y(s)$ *for all* x*,* $y \in \mathcal{A}_M$*,* $\lambda \in [0,1]$*, and* $s \in S_M$ *are also functions from* S_M *to* $\Delta(Z)$*.*

3 Axioms

In this section, we state the axioms imposed on the preference relation $\succsim$. The first axiom is standard.

Axiom 1 (Weak Order) $\succsim$ is complete and transitive.

The second axiom is also standard, except that it is restricted to menus of the same cardinalities.

Axiom 2 (Archimedian Continuity for Menus of the Same Cardinalities) For all $M = 1, 2, \cdots$ and x, y, $z \in \mathcal{A}_{\mathcal{M}}$, if $x \succ y \succ z$, then there exists λ, $\lambda' \in (0, 1)$ such that $\lambda x + (1-\lambda)z \succ y \succ \lambda' x + (1 - \lambda')z$.

Intuitively, we need the restriction because the evaluation of a menu depends on the cardinality of mental states and eventually, the cardinality of the menu, as we will indicate. If a menu is mixed with a menu of a different (typically, larger) cardinality, the mixture may have a different cardinality and be evaluated differently from the original menu, which is a source of discontinuity.

The next two are the key axioms of this paper. First, we require the independence axiom only for mixtures with singleton menus.

Axiom 3 (Singleton Independence (S-Independence)) For x, $y \in \mathcal{A}$, if $x \succsim y$ if and only if $\lambda x + (1 - \lambda)\{\beta\} \succsim \lambda y + (1 - \lambda)\{\beta\}$ for all $\lambda \in [0, 1]$ and $\beta \in \Delta(Z)$.

The intuition behind this axiom is as follows. Let x, y, and z be menus and $\lambda \in [0, 1]$. If the preference ranking between $\lambda x + (1 - \lambda)z$ and $\lambda y + (1 - \lambda)z$ differs from the ranking between x and y, the only reason is that mixing z affects the evaluations of x and y differently. However, if z is a singleton, i.e., $z = \{\beta\}$ for some $\beta \in \Delta(Z)$, then each alternative in x and y is mixed by the same alternative β in $\lambda x + (1 - \lambda)\{\beta\}$ and $\lambda y + (1 - \lambda)\{\beta\}$, which uniformly changes the evaluations of alternatives. Thus, the ranking between x and y and the ranking between $\lambda x + (1 - \lambda)\{\beta\}$ and $\lambda y + (1 - \lambda)\{\beta\}$ should be consistent.

This axiom is similar to C-independence by Gilboa and Schmeidler (1989), Set S-independence by Olszewski (2007), the independence of degenerated decisions (IDD) by Ergin and Sarver (2010). In particular, IDD adapted to correlated mixtures of menus is a weakening of S-independence (which we might refer to as weak S-independence). If we impose the weaker axiom instead of S-independence, we will obtain a more general representation similar to a costly contemplation model by Ergin and Sarver, wherein choosing the optimal probability measure over subjective states involves positive costs. However, we prefer to impose S-independence rather than the weaker axiom, since that derives a sharper result regarding the relations between preferences over menus of the different cardinalities (which we indicate in Section 5).

The next axiom requires that the DM prefer menus to mixtures of the menus with inferior menus.

Axiom 4 (Aversion to Mixtures of Inferior Menus, AMI) For all $M = 1, 2, \cdots$, x, $y \in \mathcal{A}_{\mathcal{M}}$, if $x \succsim y$, then $x \succsim \lambda x + (1-\lambda)y$ for all $\lambda \in [0, 1]$.

To understand the intuition behind this axiom, first consider exact maximization, i.e., the ex ante optimal alternative is chosen from a menu with probability 1, as the benchmark. Let alternatives $\bar{\beta}$ and $\bar{\gamma}$ be of the highest commitment ranking in menus x and y, respectively. If $\bar{\beta} \succ \bar{\gamma}$, it is natural to assume that $x \succ \lambda x + (1-\lambda)y$ since $\bar{\beta}$ is the best alternative in the first term, while $\bar{\beta}$ is mixed with an inferior alternative in the second term.

On the other hand, suppose that $\bar{\beta} \sim \bar{\gamma}$. If the DM performs exact maximization, we have $x \sim y$ by construction. In this case, mixing y with x is innocuous since $\lambda x + (1-\lambda)y$ is evaluated only by the mixture of best alternatives in x and y, i.e., $\lambda\bar{\beta}+(1-\lambda)\bar{\gamma} \sim \bar{\beta}$ under S-independence, concluding that $x \sim \lambda x + (1-\lambda)y$.

However, if the DM possibly chooses suboptimal alternatives from a menu with positive probability, we may have $x \succ y$ because suboptimal alternatives in y worsen the evaluation of the menu more than those in x. In this case, we naturally assume that mixing with y lowers the evaluation of x, and thus we have $x \succ \lambda x + (1-\lambda)y$, which is the implication of the axiom.

Axiom 4 may look similar to Ergin and Sarver's (2010) aversion to contingent planning (ACP). However, we apply a different terminology since our axiom has a different interpretation from theirs, because of correlated mixtures: Ergin and Sarver argue that making contingent plans for the mixture of menus x and y is more costly than making contingent plans for x and y themselves, and that is the reason for x to be preferred to $\lambda x+(1-\lambda)y$. However, this argument relies on the assumption of *uncorrelated* mixtures, which makes the cardinality of $\lambda x + (1-\lambda)y$ much greater than those in x and y (note that $|\lambda x + (1-\lambda)y| = |x||y|$ in this case). Since we assume correlated mixtures in this paper, which make the cardinality of $\lambda x + (1-\lambda)y$ is at most as large as the largest cardinality between x and y, the interpretation is not relevant to our axiom.

Note that the axiom is imposed only on menus of a common cardinality, since imposing it on menus of different cardinalities is too restrictive. For example, if $|x| < |y|$ for x, $y \in \mathcal{A}$, the cardinality of $\lambda x + (1-\lambda)y$ equals $|y|$. Then, x may be evaluated differently from $\lambda x + (1-\lambda)y$ as we discussed for Axiom 2, and there is no reasonable way to associate the evaluation of the former with that of the latter.

Finally, the following two axioms are technical ones imposed on $\succsim$.

Axiom 5 (Mixture-Wise Monotonicity) Let x and y be menus. If $\{\beta\} \succsim \{\gamma\}$ for all $\beta \in x$ and $\gamma \in y$ such that $\rho_{x,y}(\beta, \gamma) = 1$, then $x \succsim y$.

Axiom 6 (Nondegeneracy) There exist menus x, y such that $x \succ y$.

Intuitively, Axiom 5 says that menu x is preferred to y if the alternative chosen from x is better than the one chosen from y (with respect to the commitment ranking) in each mental state. On the other hand, Axiom 6 assures that not all menus are indifferent.

4 Representation Theorem

Now, we are in the position to state our representation theorem. First, we define an ASC representation, which is the main focus of this paper. We say that a functions $W : \mathcal{A} \to \Re$ *represents* $\succsim$ if $x \succsim y$ whenever $W(x) \geq W(y)$.

Definition 3 (Anticipated Stochastic Choice (ASC) Representation) We call that $(u, \phi, S, \mathcal{M})$ is an *anticipated stochastic choice (ASC) representation of the preference* $\succsim$ if the following conditions hold:
(i) $S = \cup_{n=1}^{\infty} S_n$, where $S_n = \{s_1, \cdots, s_n\}$;
(ii) $\mathcal{M} = \cup_{n=1}^{\infty} \mathcal{M}_n$, where $\mathcal{M}_n \subseteq \Delta(S_n)$ is convex and closed for $n = 1, 2, \cdots$;
(iii) $V : \mathcal{A} \to \Re$ is a function that represents $\succsim$ such that

$$V(x) = \max_{\mu \in \mathcal{M}_{|x|}} \int_{S_{|x|}} u(\phi(x, s)) d\mu(s), \tag{2}$$

where $u : \Delta(Z) \to \Re$ is an affine function and $\phi : \mathcal{A} \times S \to \Delta(Z)$ is a function, Further, the restriction of ϕ on $\{x\} \times S_{|x|}$, i.e., $\phi_x(\cdot) \equiv \phi(x, \cdot)$, is a bijection.

Furthermore, we say that an ASC representation $(u, \phi, S, \mathcal{M})$ of $\succsim$ is *essentially unique* if the following conditions hold for all ASC representations $(u', \phi', S', \mathcal{M}')$ of $\succsim$:
(iv) u is unique up to a positive affine transformation, i.e., there exists $a > 0$ and $b \in \Re$ such that $u'(\beta) = au(\beta) + b$ for all $\beta \in \Delta(Z)$;
(v) there exists a bijection $\eta : S \to S'$ such that $\phi'(x, \eta(s)) = \phi(x, s)$ for all $x \in \mathcal{A}$ and $s \in S$;
(vi) $\mathcal{M} = \{\mu \circ \eta \in \Delta(S) : \mu' \in \mathcal{M}\}$.

This representation is interpreted as follows. Conditions (i) and (ii) define the sets of mental states S and state-generating functions, i.e., probability distributions over the sets of mental states, $\mathcal{M}$. Condition (iii) denotes a representation V of $\succsim$, where a choice function ϕ yields an alternative in a given menu x, depending on the mental state $s \in S_{|x|}$. For a menu x, a state-generating function $\mu \in \mathcal{M}_{|x|}$ together with ϕ derives a probabiilty distribution over alternatives in a given menu x, which is interpreted as a stochastic choice function, i.e., a probability distribution over alternatives in x.[8] Fully anticipating this, the DM chooses the optimal state-generating function μ to maximize the expected utility with the utility function u.[9] Note that the bijectiveness of ϕ_x guarantees the minimality of the set of mental states S_n for each n, i.e., if $S'_n \supsetneq S_n$, then S'_n is never the set of mental states. On the other hand, uniqueness condition (iv) is straightforward. The implication of (v) is that the sets of mental states that induce the same ex post choice after renumbering are considered equivalent, and (vi) implies that after the renumbering, the sets of state-generating functions are equivalent.

Now, we state the main theorem of this paper.

Theorem 1 (Main Theorem) *Under Assumption 1, $\succsim$ satisfies Axioms 1-6 if and only if $\succsim$ admits an ASC representation $(u, \phi, S, \mathcal{M})$. Furthermore, the ASC representation is essentially unique.*

The proof is in the Appendix. A sketch of the proof is as follows: first, weak order, continuity, and S-independence imply that there exists a von Neumann-Morgenstern function u that represents the commitment ranking $\succsim$. Without loss of generality, u is normalized to $K \subseteq \Re$.

Next, fix a finite number M of alternatives in each menu. Since Assumption 1 assures that a mental state can be associated with an alternative of an arbitrary commitment ranking, there exists a set of mental states S_M such that the set of menus x and the set of functions $\phi_x : S_M \to x$ are isomorphic. Moreover, by the definition of correlated mixtures that we employ, the set of composite functions $u \circ \phi_x : S_M \to K$ of utility functions u and functions ϕ_x constructs a linear space, wherein addition and scalar multiplication are derived from the perfectly correlated mixture coefficient.

Hence, over functions $\phi_x : S_M \to x$, there exists a preference relation that satisfies the counterparts of Axioms 1-6, which are similar to those imposed by Gilboa and Schmeidler (1989). (Note that the counterpart of AMI has the inverse implication of Gilboa and Schmeidler's uncertainty aversion, which makes our representation the maximum of expected utility as opposed to the minimum.) Thus, an argument similar to theirs derives the desired preference representation for the fixed M. Finally, since the same argument holds for all M, the desired representation is obtained for any menu x. The uniqueness result is similar to that of Gilboa and Schmeidler (1989).

[8] The stochastic choice function employed here is standard in the literature on stochastic choice (e.g., McFadden and Richter 1991) except that mental states s, which determine the choice of an alternative from a menu, are explicit in an ASC representation, while they are implicit in the literature.

[9] We can easily derive a representation in which a state-generating function is chosen so that the expected utility is minimized from reversing the implication of AIM. The obtained representation implies that ex post choice of an alternative is made by nature, which makes the DM ambiguity averse. We will discuss more on the difference in implication between the modified representation and an ASC representation in Subsection 8.2.

A few remarks are in order. First, the (essential) uniqueness of S and $\mathcal{M}$ in an ASC representation renders the interpersonal comparison of $\mathcal{M}$ meaningful: a larger set $\mathcal{M}$ of state-generating functions provides the DM more freedom to control her future self, which can be interpreted as more rationality. Conversely, a smaller $\mathcal{M}$ restricts the DM's ability to control her future self, thus she prefers commitment to singleton menus more, to avoid choosing suboptimal alternatives. We will formalize this discussion in Subsection 6.2.

Second, an ASC representation generally derives probability distributions over the alternatives of a menu that do *not* generally satisfy extremeness, i.e., ASC representations permit non-extreme points to be chosen with positive probability. For example, let $Z \subset \Re^2$ be $\{b_1, b_2, b_3, b_4\}$, where $b_1 = (0, 0)$, $b_2 = (0, 1)$, $b_3 = (1, 0)$, and $b_4 = (1/3, 1/3)$. Then, by assuming that all $\mu \in \mathcal{M}_4$ have full supports, b_4 is chosen from $x = \{b_1, b_2, b_3, b_4\}$ (where b_i denote degenerated lotteries for i = 1, 2, 3, and 4) with positive probability. This property renders an ASC representation more tractable than the random utility models such as Dekel, Lipman, and Rustichini (2001), Dekel and Lipman (2010), and Ergin and Sarver (2010) as a stochastic choice model, as we discussed in Subsection 1.2.

Finally, "on-path" stochastic choice functions, i.e., probability distributions over menus induced by the optimal state-generating functions and the choice function, generally violate regularity (Luce and Suppes 1965) and Weak Axiom of Revealed Stochastic Preference (WARSP, Bandyopadhyay et al. 1999).[10] This property provides flexibility to an ASC representation to accommodate some choice anomalies as we will demonstrate in Section 8.

We conclude this section by discussing four special cases of an ASC representation. The first is $\mathcal{M}_n = \Delta(S_n)$ for $n = 1, 2, \cdots$. In this case the DM can exactly attain the optimality with probability 1 since she chooses the optimal state-generating function among all the probability measures over S_n. We will demonstrate in Subsection 6.1 that monotonicity implies (and is implied by) this special case.

The second is $\mathcal{M}_n = \{\mu_n\}$ for some probability measures $\mu_n \in \Delta(S_n)$ with $n = 1, 2, \cdots$. This corresponds to the pure stochastic choice case, i.e., the DM chooses alternatives with probability specified by μ_n. In particular, if μ_n is such that $\mu_n(\{s_1\}) = 1$ and $\mu_n(\{s_i\}) = 0$ for $i \neq 1$, only the alternative corresponding to s_1 is chosen with probability 1 regardless of its commitment ranking, which is interpreted that choice is so inadvertent that only the most salient alternative is chosen with positive probability. We will discuss these special cases in Section 7.

[10]Our approach shares the same spirit with Noor and Takeoka (2009, 2010), who consider a preference for menus that violates independence or WARP (Weak Axiom of Revealed Preference). However, they focus on deterministic choice of an alternative from a menu, whereas we focus on stochastic choice (they examine WARP while we examine WARSP).

The third special case of interest is $\mathcal{M}_n = \{(1-\epsilon)\mu + \epsilon\mu_n : \mu \in \Delta(S_n)\}$ for some state-generating function $\mu_n \in \Delta(S_n)$ and $n = 1, 2, \cdots$. In this case, all the state-generating functions are available with probability $1-\epsilon$ and only a purely stochastic choice is available with probability ϵ. Eventually, we obtain a trembling-hand representation (Selten 1975)

$$V(x) = (1-\epsilon)\max_{\beta\in x} u(\beta) + \epsilon \int_{S_{|x|}} u(\phi(x,s))d\mu_{|x|}(s), \tag{3}$$

which implies that the DM chooses the best alternative with probability $1-\epsilon$ and arbitrary alternatives with probability ϵ.[11][12]

Finally, choice under limited consideration sets (Wright and Barbour 1977; Masatlioglu, Nakajima, and Ozbay 2010) can be also considered as a special case of an ASC representation. For each $n = 1, 2, \cdots$, let $Q_n \subsetneq S_n$ and $\mathcal{M}_n = \{\mu \in \Delta(S_n) : \mathrm{supp}(\mu) \subseteq Q_n\}$ ($\mathrm{supp}(\mu)$ denotes the support of μ). Then, the ASC representation becomes $V(x) = \max_{\beta\in\phi(x,Q_{|x|})} u(\beta)$. In other words, the DM chooses the best alternative from the ones corresponding to mental states in Q_n, and the alternatives corresponding to mental states in $S_n \setminus Q_n$ will be never "considered," irrespectively of their utility. Compared to a revealed attention model by Masatlioglu, Nakajima, and Ozbay (2010), the ASC representation permits stochastic choice as well as deterministic choice, whereas the "consideration sets" $\phi(x, Q_n)$ depend on the correspondences between mental states and alternatives.

5 Relations between Preferences for Menus of Different Cardinalities

In the previous section, we have not specified relations between the sets of mental states associated with menus with different cardinalities. In other words, if the cardinalities of menus x and y are different, there is not necessarily a correspondence between mental states associated with x and y. On the other hand, the following analysis relates the sets of mental states by imposing the following axiom.

[11]For the sake of simplicity, we do not provide a formal axiomatization for a trembling-hand ASC representation. However, the axiomatization will be easily obtained by exploiting an axiomatization of ϵ-contamination (for example, Nishimura and Ozaki 2006), and the similarity between ASC and maxmin expected utility representations.

[12]This representation may remind the reader of Chatterjee and Krishna (2009), who consider the DM that maximizes the ex ante utility with probability $1-\epsilon$ and the "alter-ego" utility with probability ϵ. In particular, if we accept the second interpretation in Subsection 2.1, i.e., ex post choices in an ASC representation are made to maximize the ex post utility given mental states, (3) implies that the DM maximizes the ex ante utility with probability $1-\epsilon$ and the multiple "alter-ego" utility, which depends on mental states, with probability ϵ.

Unlike their study, however, the ex post utility function in this interpretation is not generally a von Neumann-Morgenstern utility function. Accordingly, an ASC representation does not satisfy extremeness as we mentioned, while their model satisfies it.

Axiom 7 (Indifference to Addition of Commitment Equivalent, IAC) For $x \in \mathcal{A}$, if $x \sim \{\bar{\beta}\}$ and $\bar{\beta}$ is a mixture-invariant addition to x, then $x \cup \{\bar{\beta}\} \sim \{\bar{\beta}\}$.

Suppose that the menu x is indifferent to a singleton menu $\{\bar{\beta}\}$ (which we refer to as the *commitment equivalent of* x). Then, it is reasonable to assume that $x \cup \{\bar{\beta}\}$ is also indifferent to $\{\bar{\beta}\}$ since the addition of $\bar{\beta}$ does not change the existing correspondences between alternatives in x and mental states, which explains the implication of the axiom.

Now, we consider the following relation between the sets of state-generating functions.

Definition 4 (Set of State-Generating Functions Derived from the Bayes' Rule) Let $\mathcal{M}_{n+k}$ be the set of state-generating functions over S_{n+k} in an ASC representation $(u, \phi, S, \mathcal{M})$. We refer to $\mathcal{M}^*_{n,k}$ as the *set of state-generating functions over* S_n *derived from the Bayes rule* if the following condition is satisfied:

$$\mathcal{M}^*_{n,k} = \{\mu(S_n \cap \cdot)/\mu(S_n) | \mu \in \mathcal{M}_{n+k}, \mu(S_n) > 0\}. \tag{4}$$

In other words, $\mathcal{M}^*_{n,k}$ consists of state-generating functions derived from the Bayes' rule applied to each state-generating function in $\mathcal{M}_{n+k}$, while excluding the ones that assign zero probability to S_n. If the sets of state-generating functions are singletons, then $\mathcal{M}^*_{n,k}$ are simply derived from the Bayes' rule applied to the singletons, which is compatible with Luce (1959) (we will formally characterize this case in Section 7).

Now, the following theorem indicates that under Axiom 7, $\mathcal{M}_n$ includes state-generating functions that are derived from the Bayes' rule applied to each state-generating function in $\mathcal{M}_{n+k}$.

Theorem 2 *Assume that $\succsim$ has an ASC representation $(u, \phi, S, \mathcal{M})$. Then, IAC implies the following statements for all n, $k \in N$.*
(i) $\mathcal{M}^*_{n,k} \subseteq \mathcal{M}_n$, and
(ii) $\mathcal{M}^*_{n,k} = \mathcal{M}_n$ if $\mu(S_n) > 0$ for all $\mu \in \mathcal{M}_{n+k}$.

The proof is relegated to the Appendix. To see the reason that we need the condition that $\mu(S_n) > 0$ for all $\mu \in \mathcal{M}_{n+k}$ to obtain statement (ii), note that a key step of the proof is the equality in (10), i.e., $\max_{\mu \in \mathcal{M}^*_{n,k}} \int_{S_n} u \circ \phi(x, s) d\mu = \max_{\mu \in \mathcal{M}_{n+k}} \int_{S_n} \frac{u \circ \phi(x,s)}{\mu(S_n)} d\mu$, the right-hand side of which is not well-defined if there exists $\mu \in \Delta(S_{n+k})$ such that $\mu(S_n) = 0$.

Note also that dropping the condition in Axiom 7 that $\bar{\beta}$ is a mixture-invariant addition is too restrictive to obtain a general ASC representation: for example, let $x = \{\beta_1, \beta_2\}$ and $x^i = x \cup \{\bar{\beta}^i\}$ with $\bar{\beta}^i \in \Delta(Z)$ such that $\rho_{x^i, x_3^*}(\bar{\beta}^i, \beta_i^*) = 1$ for $i = 1$, 2, and 3. The strengthened version of Axiom 7 implies that x^1, x^2, and x^3 are all indifferent, while the original statement only implies that x^3 is indifferent to $\{\bar{\beta}\}$. However, since $\bar{\beta}^1$, $\bar{\beta}^2$, and $\bar{\beta}^3$ have the same commitment ranking, yet mixed with different β_i^*, the indifference between x^1, x^2, and x^3 implies that the alternative of the same commitment ranking is chosen from each menu with the same probability (e.g., $\bar{\beta}^i$ is chosen from x^i with the same probability for $i = 1$, 2, and 3). The only ASC representation that satisfies the last condition is such that the set of state-generating functions for each cardinality of menus is a singleton that gives the same probability to all mental states.

6 Monotonicity, Preferences for Commitment, and Set Betweenness

In this section, we explore an important implication of an ASC representation regarding the size of menus – whether the DM prefers flexibility or commitment – which is often examined in the literature.

6.1 Monotonicity

In this subsection, we discuss monotonicity, i.e., a preference for larger menus, which is interpreted as a preference for flexibility. The following axiom is often discussed in the literature on menu choice (e.g., Kreps 1979, 1992; Dekel, Lipman, and Rustichini 2001; Nehring 1999; Ergin and Sarver 2010).

Axiom 8 (Monotonicity) For all x, $y \in \mathcal{A}$, if $x \supseteq y$, then $x \succsim y$.

The following theorem indicates that an ASC does not generally satisfy monotonicity unless the sets of state-generating functions are equivalent to the sets of all probability measures over mental states.

Theorem 3 *Suppose that $\succsim$ has an ASC representation $(u, \phi, S, \mathcal{M})$. Then, the following statements are equivalent.*

(a) $\succsim$ satisfies monotonicity
(b) $\mathcal{M}_n = \Delta(S_n)$ for $n = 1, 2, \cdots$.

This theorem indicates that there is no nontrivial intersection between an ASC representation and the existing studies on a preference for flexibility à la Dekel, Lipman, and Rustichini (2001). Namely, whenever we require monotonicity, an ASC representation reduces to a usual utility maximization model with a state-independent utility function.

6.2 Preferences for Commitment

In this subsection, we investigate the other extreme, that is, a preference for commitment.

Definition 5 (Preferences for Commitment to Singleton Menus) (1) For $x \in \mathcal{A}$, preference $\succsim$ exhibits *a preference for commitment to a singleton menu at* x if there exists $\beta \in x$ such that $\{\beta\} \succ x$.

(2) *Preference* $\succsim_2$ *exhibits a preference for commitment to a singleton menu more than preference* $\succsim_1$ if for all $x \in \mathcal{A}$, there exists $\beta \in x$ such that

$$\{\beta\} \succsim_1 x \Rightarrow \{\beta\} \succsim_2 x. \tag{5}$$

Unlike Gul and Pesendorfer (2001), we restrict our attention to commitment to singleton menus, which implies that a preference for commitment to a singleton menu is stronger than a preference for commitment in their sense. Dekel and Lipman (2010) also focus on a preference for commitment to a singleton menu and Ergin and Sarver (2010) employ a similar condition to compare contemplation costs. The following observation indicates that an ASC preference exhibits a preference for commitment to a singleton menu, which also implies that it exhibits a preference for commitment in Gul and Pesendorfer's sense.

Observation 1 *Suppose that* $\succsim$ *has an ASC representation* $(u, \phi, S, \mathcal{M})$ *such that* $\mathcal{M} \subsetneq \cup_{n=1}^{\infty} \Delta(S_n)$. *Then, there exists* $x \in \mathcal{A}$ *with* $|x| \geq 2$ *such that* $\succsim$ *exhibits preference for commitment to a singleton menu at* x.

To see why this implication holds, let β be the alternative of the highest commitment ranking in x. Then, it is straightforward from the definition of an ASC representation that $\{\beta\} \succ x$ unless $\mathcal{M}_{|x|} = \Delta(S_{|x|})$.

As is noted in Section 4, the size of the set of state-generating functions $\mathcal{M}_n$ can be interpreted as an index for the rationality of the DM. We provide a behavioral foundation for this interpretation.

Theorem 4 *Suppose that the* DM_i*'s preference* $\succsim_i$ *has ASC representations* $(u, S, \mathcal{M}^i)$ *for* i $= 1, 2$. *Then,* $\succsim_2$ *exhibits a preference for commitment to a singleton menu more than* $\succsim_1$ *if and only if* $\mathcal{M}^1 \supseteq \mathcal{M}^2$.

The proof can be found in the Appendix. Since the set of state-generating functions in our representation describes state-generating functions from which the DM chooses the optimal one, a smaller set of state-generating functions for DM2 than that for DM1 implies that DM2 is more likely to "mistakenly" choose suboptimal alternatives from a menu ex post than DM1. Eventually, DM2 prefers commitment more than DM1 to avoid unintentionally choosing suboptimal alternatives.

In a trembling-hand case, the magnitude of ϵ implies (and is implied by) the set inclusion condition above. Let $\mathcal{M}_n = \{(1-\epsilon)\mu+\epsilon\mu_n : \mu \in \Delta(S_n)\}$ and $\mathcal{M}'_n = \{(1-\epsilon')\mu+\epsilon'\mu_n : \mu \in \Delta(S_n)\}$ (note that μ_n is common between $\mathcal{M}_n$ and $\mathcal{M}'_n$). Then, $\epsilon \leq \epsilon'$ if and only if $\mathcal{M}_n \supseteq \mathcal{M}'_n$.

The proof of this theorem parallels with Ghirardato and Marinacci (2002, Theorem 17), who relates an interpersonal comparison of ambiguity aversion to that of the set of priors in maxmin expected utility. We should note, however, that we have the inverse implication about the relation between the size of the set of state-generating functions and a preference for singleton menus: the contraction of the set of priors is associated with a *less* preference for certainty in their study, whereas the contraction of the set of state-generating functions is associated with a *more* preference for commitment to a singleton menu in this paper. This reflects the fact that the optimal state-generating function maximizes, instead of minimizes as in maxmin expected utility, the expected utility.

6.3 Set Betweenness

Until this point, we have discussed the two extreme case, a preference for flexibility and commitment. In this section, we explore something intermediate, i.e., set betweenness proposed by Gul and Pesendorfer (2001).

Axiom 8' (Set Betweenness) Let $x, y \in \mathcal{A}' \subseteq \mathcal{A}$. If $x \succsim y$, then $x \succsim x \cup y \succsim y$.

It is easy to see that an ASC representation satisfies betweenness if it is restricted to singleton menus. The statement is formalized as follows.

Observation 2 *Suppose that* $\succsim$ *has an ASC representation* $(u, \phi, S, \mathcal{M})$. *Then, it satisfies set betweenness if* $\mathcal{A}' = \mathcal{A}_1$.

Proof For singleton menus $x = \{\beta\}$, $y = \{\gamma\}$, β, $\gamma \in \Delta(Z)$, the utilities of x and y are the (expected) utilities of choosing β and γ with probability 1, respectively. Accordingly, the utility of $x \cup y = \{\beta, \beta'\}$ is the weighted sum of those of β and γ, whatever state-generating function is chosen from the set, which leads to $x \succsim x \cup y \succsim y$ whenever $x \succsim y$. **Q.E.D.**

The reader may suspect that set betweenness holds for non-singleton menus. However, it is not generally true in an ASC representation. Consider the following example: let $x = \{\beta_1\}$, $y = \{\beta_2, \beta_3\}$ such that $\beta_3 \succ \beta_1 \succ \beta_2$. Suppose that $\mathcal{M}_2 = \{\mu\}$ such that $\mu \in \Delta(S_2)$ and $\mu(s_1) = 1$, $\mathcal{M}_3 = \{\mu'\}$ such that $\mu' \in \Delta(S_3)$ and $\mu'(s_3) = 1$, and $\phi(x \cup y, s_i) = \beta_i$ for $i = 1$, 2, and 3. Then, we obtain $x \succ y$ whereas $x \cup y \succ x$, which violates set betweenness.

To see the reason for the violation, notice the following two effects of the increase in the cardinality of a menu, which results from combining menus x and y. The first is reducing the probability of choosing the best alternative from the combined menu, given that the state-generating function remains unchanged. The second is altering the set of state-generating functions from which the optimal state-generating function is chosen. Singleton menus eliminate the second effect since they derive singleton sets of state-generating functions for which the only alternatives are chosen with probability 1, while they preserve the first.

Note that set betweenness is generally violated for non-singleton menus even under IAC: let $\mathcal{M}'_3 = \{\mu' \in \Delta(S_3) : \mu'(s_1) + \mu'(s_3) = 1\}$, which clearly includes $\mathcal{M}_3$ in the example above. Then, we have $\mathcal{M}^*_{2,1} = \mathcal{M}_2$. Hence, it follows from Theorem 2 that an ASC representation with $\mathcal{M}_2$ and $\mathcal{M}'_3$ satisfies IAC, yet violates set betweenness.

7 Pure Stochastic Choice

In this section, we investigate a special case of ASC representation, in which the set of state-generating functions is a singleton, which corresponds to Luce's (1959) choice axiom approach.

Definition 6 (Pure Stochastic Choice) A *pure stochastic choice representation* is an ASC representation $(u, \phi, S, \mathcal{M})$ such that $\mathcal{M} = \cup_n^\infty \{\mu_n\}$ for $\mu_n \in \Delta(S_n)$, $n = 1, 2, \cdots$.

Next, we consider the following axiom, which is a strengthening of both S-independence and AMI.

Axiom 3' (Independence for Menus of the Same Cardinalities) For all $M = 1, 2, \cdots$, x, y, $z \in \mathcal{A}_{\mathcal{M}}$, and $\lambda \in [0, 1]$, $x \succsim y$ if and only if $\lambda x + (1 - \lambda)z \succsim \lambda y + (1 - \lambda)z$.

This axiom implies that for all menus z of the same cardinality as x and y, the ranking between menus x and y never changes before and after they are mixed with z. Note that the reason that we restrict our attention to menus of the same cardinality is to exclude the change in the evaluation of menus resulted from the change of cardinality.

From a similarity of ASC representation to maxmin expected utility theory, we expect that imposing the axiom derives a pure stochastic choice representation. The following theorem indicates that this conjecture is correct.

Theorem 5 *Under Assumption 1, preference $\succsim$ satisfies Axioms 1, 2, 3', 5, and 6 if and only if $\succsim$ admits a pure stochastic choice representation.*

The proof is in the Appendix. Note that an ASC preference does not satisfy Axiom 3' unless it is pure stochastic choice. For example, consider the exact maximization case, i.e., $\mathcal{M} = \cup_{n=1}^{\infty}\Delta(S_n)$, and let $x = \{\beta_1, \beta_2\}$, $y = \{\beta_1', \beta_2'\}$, and $z = \{\gamma_1, \gamma_2\}$ with $\phi(x, s_i) = \beta_i$, $\phi(y, s_i) = \beta_i'$, $\phi(z, s_i) = \gamma_i$ for $i = 1, 2$. Then, $x \succsim y$ implies that the best (with respect to the commitment ranking) alternative β_i of x is preferred to the best alternative β_j' of y. Now, consider mixtures $\lambda x + (1-\lambda)z = \{\lambda\beta_1 + (1-\lambda)\gamma_1, \lambda\beta_2 + (1-\lambda)\gamma_2\}$ and $\lambda y + (1-\lambda)z = \{\lambda\beta_1' + (1-\lambda)\gamma_1, \lambda\beta_2' + (1-\lambda)\gamma_2\}$. Then, $\lambda x + (1-\lambda)z \succsim \lambda y + (1-\lambda)z$ with exact maximization implies that the best alternative $\lambda\beta_k + (1-\lambda)\gamma_k$ of $\lambda x + (1-\lambda)z$ is preferred to the best alternative $\lambda\beta_l' + (1-\lambda)\gamma_l$ of $\lambda y + (1-\lambda)z$. However, it is not necessarily the case that $i = k$ and $j = l$ since mixing γ_k and γ_l may change the commitment ranking among alternatives.

It is straightforward from Theorem 5 that together with IAC, state-generating functions in a pure stochastic choice representation satisfy the Bayes' rule.

Corollary 1 *Suppose that $\succsim$ has a pure stochastic choice representation $(u, \phi, S, \mathcal{M})$. Then, IAC holds if and only if*

$$\mu_n(E) = \frac{\mu_{n+k}(E \cap S_n)}{\mu_{n+k}(S_n)}$$

for all n, $k = 1, 2, \cdots$, and $E \subseteq S$.

In a pure stochastic choice case, we can consider several variations. To obtain the results, we assume the existence of a complete and transitive salience ranking $\succsim^*$.

Definition 7 (Salience Ranking) We refer to a complete and transitive order $\succsim^*$ as the *salience ranking* if for all $x \in \mathcal{A}$ and β, $\beta' \in x$ such that $\phi(x, s_i) = \beta$ and $\phi(x, s_j) = \beta'$, $\beta \succsim^* \beta'$ implies that $i \leq j$. Moreover, $\beta \sim^* \beta'$ implies that $\beta = \beta'$.

A possible interpretation of $\succsim^*$ is such that $\beta_i \succsim^* \beta_j$ implies that "β_i is (weakly) more salient (thus, more prone to inadvertent choice) than β_j."[13] Note that unless otherwise stated, the statement does not imply that β_i is chosen with higher probability than β_j from x. In particular, unlike a preference order, $\beta \succ^* \beta'$ for all $\beta' \in x$ does not necessarily imply that β is chosen from x with probability 1. Conversely, if there exists a salience ranking $\succsim^*$, a perfectly correlated mixture coefficient can be derived from $\succsim^*$ as follows: for all $x = \{\beta_1, \cdots, \beta_n\}$ and $y = \{\gamma_1, \cdots, \gamma_m\}$ such that $\beta_1 \succ^* \cdots \succ^* \beta_n$, $\gamma_1 \succ^* \cdots \succ^* \gamma_m$, and $n \geq m$, define $\rho_{x,y}(\beta_i, \gamma_i) = 1$ for $i \leq m$ and $\rho_{x,y}(\beta_i, \gamma_m) = 1$ for $i > m$.

Now we consider the following axiom.

Axiom 9 (Preferences for Improvements in More Salient Alternatives) For n, i, $j \in N$, $i \leq j$, $\beta_1, \succ^* \cdots \succ^* \beta_n$, $\beta \sim^* \beta_i$, $\beta' \sim^* \beta_j$, and $\beta \sim \beta' \succ \beta_1 \sim \cdots \sim \beta_n$. Then, $\{\beta_1, \cdots, \beta_{i-1}, \beta, \beta_{i+1}, \cdots, \beta_n\} \succsim \{\beta_1, \cdots, \beta_{j-1}, \beta', \beta_{j+1}, \cdots, \beta_n\}$.

This axiom states that the higher the salience ranking of an alternative is, the larger the impact of an improvement (in terms of the commitment ranking) in the alternative on the evaluation of a menu becomes. While the previous sections do not assume a specific order between mental states, we demonstrate in the next theorem that for a pure stochastic choice representation, Axiom 9 assures monotonicity with respect to the salience ranking, i.e., alternatives associated with higher salience rankings are chosen with higher probability.

Theorem 6 *Suppose that $\succsim$ has a pure stochastic choice representation $(u, \phi, S, \mathcal{M})$. Then, preferences for improvements in more salient alternatives is satisfied if and only if $\mu(s_i) \geq \mu(s_j)$ for $i \leq j$.*

The proof is in the Appendix. Note that without assuming a pure stochastic representation or independence, Axiom 9 does not necessarily derive monotonicity with respect to the salience ranking for a general (non-pure stochastic choice) ASC representation for *all* $\mu \in \mathcal{M}$. The reason is that the axiom has a bite only on the "on-path" state-generating function, which is actually chosen from the set of state-generating functions, and cannot specify the forms of other state-generating functions.

The next axiom, which is a strengthen of the previous axiom, consider the case that inadvertent choice is dominant and the ranking between menus is solely determined by the alternatives

[13]Manzini and Mariotti (2010) and Bordalo, Gennaioli, and Shleifer (2010) also consider a ranking similar to ours. Our model differs from the former in that we derive a choice probability from the menu preference rather than directly assuming it. It differs from the latter in that we focus on menu choice rather than the choice of alternatives, and stochastic choice (derived from the menu choice) rather than deterministic choice.

of the highest salience ranking within menus, independently of its commitment ranking.

Axiom 9' (Salience Ranking Dominance) Let $x = \{\beta_1, \cdots, \beta_m\}$ and $y = \{\beta'_1, \cdots, \beta'_n\}$ with $\beta_1 \succ^* \cdots \succ^* \beta_m$ and $\beta'_1 \succ^* \cdots \succ^* \beta'_n$. Then, $x \succsim y$ if and only if $\beta_1 \succsim \beta'_1$.

The next theorem characterizes a pure stochastic choice representation that satisfies the axiom.

Theorem 7 *Suppose that $\succsim$ has an ASC representation $(u, \phi, S, \mathcal{M})$. Then, salience ranking dominance is satisfied if and only if $\succsim$ has a pure stochastic choice representation such that $\mu_n(s_1) = 1$ and $\mu_n(s_i) = 0$ for $i \neq 1$, for $n = 1, 2, \cdots$.*

The proof is in the Appendix. Intuitively, since the ranking among menus is solely determined by the alternative of the highest salience ranking, positive probability is assigned only to the alternative and never assigned to other alternatives.

Note that this theorem only assumes an ASC representation, not a pure stochastic choice representation, which contrasts with Theorem 6. In other words, Axiom 9' is strong enough to obtain a (special case of) pure stochastic choice representation without the help of independence.

This special case roughly corresponds to Strotz (1955) and an overwhelming temptation representation by Gul and Pesendorfer (2001), wherein the ex post preference dominates the choice of an alternative from a menu, yet the desirable alternative is determined by the ex ante (or commitment) preference.

8 Discussion

8.1 Anomalies

In this subsection, we demonstrate that ASC representations accommodate choice anomalies often reported by the literature. In the following, we denote $c(x) = (\beta_1, \cdots, \beta_n)$ where $x = \{\beta_1, \cdots, \beta_n\}$, to imply that $\phi(x, s_i) = \beta_i$ for $i = 1, \cdots, n$ for simplicity.

8.1.1 Attraction Effect

The attraction effect is such that adding an unchosen alternative to the existing menu changes the choice of an alternative (for examples in specific contexts, see, e.g., Huber et al. 1982).

Typically, let β, γ, and $\delta_\beta \in \Delta(Z)$. Suppose that γ is preferred to β, but the addition of δ_β, which is dominated by β (and often referred to as a "decoy") attracts the DM's attention to β. Then, the DM chooses γ from $x = \{\beta, \gamma\}$ and chooses β from $y = \{\beta, \gamma, \delta_\beta\}$ with the highest probabilities (or often, probability 1). This example clearly violates WAR(S)P since the addition of δ_β changes the alternative chosen with the highest probability (or probability 1 in the case of WARP) from γ to β.

It is easy to confirm that this behavioral pattern can be explained if we assume that $\gamma \succ \beta \succ \delta_\beta$, $c(y) = (\beta, \delta_\beta, \gamma)$, $\mathcal{M}_2 = \{\mu \in \Delta(S_2) : \mu(s_i) \leq 1 - \epsilon \text{ for } i = 1, 2\}$, and $\mathcal{M}_3 = \{\mu \in \Delta(S_3) : \mu(s_3) \leq \epsilon'\}$ for sufficiently small ϵ, ϵ'. In other words, despite γ has the highest commitment ranking among the three alternatives (and is chosen from x with probability $1-\epsilon$), it is chosen from y with very low probability (equal to ϵ'), since the addition of δ_β drives γ out to a "blind spot."

8.1.2 Cyclical Choice

Let β, γ, $\delta \in \Delta(Z)$. Suppose that $c(\{\beta, \gamma\}) = (\beta, \gamma)$, $c(\{\gamma, \delta\}) = (\gamma, \delta)$, $c(\{\beta, \delta\}) = (\delta, \beta)$ and $\mathcal{M}_2 = \{\mu \in \Delta(S_2) : \mu(s_1) = 1 - \epsilon\}$ for a sufficiently small ϵ.

Then, β is chosen from $\{\beta, \gamma\}$, γ is chosen from $\{\gamma, \delta\}$, and δ is chosen from $\{\gamma, \delta\}$ with probability $1 - \epsilon$, respectively. Namely, the more salient alternative (the alternative associated with s_1) in each pair is chosen with high probability, irrespectively of the commitment ranking. This example indicates that in an ASC representation, $\mathcal{M}$, the set of state-generating functions, and ϕ (or equivalently, c), the correspondences between mental states and alternatives, may have more impact on the choice of alternatives than the commitment ranking.

Note that there does not exist a (transitive) salience ranking $\succsim^*$ that implements the conditions above, since the correspondences between alternatives and states themselves consist a cycle.

8.1.3 Allais Paradox

A typical form of Allais Paradox is as follows: the DM prefers a degenerated lottery α' yielding \$1m with certainty to a lottery β' yielding \$5m, \$1m, and 0 with probability .1, .89, and .01, respectively, whereas she prefers a lottery β'' yielding \$5m with probability .1 to a lottery α'' yielding \$1m with probabiilty .11. Note that replacing a .89 chance of obtaining \$1m by 0 in the former pair derives the latter.

To abstract the idea, let α, β, γ $\delta \in \Delta(Z)$, and define $\alpha' = \lambda\alpha + (1-\lambda)\gamma$, $\beta' = \lambda\beta + (1-\lambda)\gamma$, $\alpha'' = \lambda\alpha + (1 - \lambda)\delta$, and $\beta'' = \lambda\beta + (1 - \lambda)\delta$ for some $\lambda \in [0, 1]$. Assume that $c(\{\alpha', \beta'\}) = (\alpha', \beta')$, $c(\{\alpha'', \beta''\}) = (\beta'', \alpha'')$, and $\mathcal{M}_2 = \{\mu \in \Delta(S_2) : \mu(s_1) \leq 1-\epsilon, \mu(s_2) \leq \epsilon'\}$ for sufficiently small ϵ and ϵ' with $\epsilon \leq \epsilon'$.

Suppose that $\alpha' \succ \beta'$, which implies that α and β are chosen from $x = \{\alpha', \beta'\}$ with probabilities $1 - \epsilon$ and ϵ, respectively. Assuming that the commitment ranking is a von Neumann-

Morgenstern utility function (which is the case with an ASC representation), we have $\alpha'' \succ \beta''$. However, the ASC representation implies that β'' and α'' are chosen from $y = \{\alpha'', \beta''\}$ with probabilities $1 - \epsilon'$ and ϵ', respectively, which explains Allais paradox.

The intuition is that α' is chosen from x with higher probability since it is preferred to β', while β'' is chosen from y with higher probability since it is more salient than α''. For example, the \$5m prize of β'' may capture the DM's eye in the second pair, which results in the choice of β'' over α'', in a way somewhat different from the way she chooses from the first pair.

8.1.4 Summary

We have shown that various choice anomalies, typically, choice reversal results can be explained by an ASC representation. The crucial aspect here is that the probability that an alternative is chosen is determined not only by the commitment ranking of the alternative but also by $\mathcal{M}$, the set of state-generating functions.

Since the unions of supports of all state-generating functions in $\mathcal{M}_n$ can be interpreted as consideration sets as we mentioned, the implications of these results are similar to those of Masatlioglu, Nakajima, and Ozbay (2010), who employ consideration sets to explain anomalies. Unlike their study, however, our model accommodate stochastic choice in addition to deterministic choice (note that the examples above all permit stochastic choice), which derives a richer implication.

8.2 Self vs. Nature

In this paper, we model the DM who chooses a state-generating function, i.e., probability measure over mental states, so that her expected utility is *maximized*. Ergin and Sarver (2010) consider a similar model (yet in a different context), whereby a probability measure over future preferences is chosen to maximize expected utility. On the other hand, Gilboa and Schmeidler (1989) modeled an ambiguity averse DM, who chooses a probability measure over the (exogenous) states of nature so that her expected utility is *minimized*, and Epstein, Marinacci, and Seo (2007) extended this approach to a subjective state space, which describes unforeseen contingencies. Similar arguments are also made by Olszewski (2007) and Ahn (2007), who identify an increase of ambiguity with an expansion of a set of alternatives.

The difference between these two approaches is summarized as whether self or nature chooses an alternative from a menu at the ex post stage: suppose that ex post choice (or preference, depending on the interpretation) depends only on internal (or psychological) states within the DM, such as cravings, absent-mindedness, and proneness to heuristics and biases. Being under the influence of such internal states, her ex post self chooses an alternative from a menu. In this case, it is reasonable to assume that the DM can choose a probability measure over states to maximize her expected utility, since she has control of her future self (at least, to a certain extent).

On the other hand, suppose that the ex post preference depends only on external states such as weather, natural disasters, and booms and slumps, which can be interpreted that nature chooses an alternative from a menu ex post. In this case, she may be pessimistic about the realizations of states since these states are out of reach of the DM's control, and she is only concerned about the worst-case scenario, i.e., the minimum of the expected utility.

Accordingly, if we reverse the implication of AIM and eventually obtain the counterpart of an ASC representation, in which a state-generating function is chosen so that the expected utility is *minimized* (the derivation of such a representation is straightforward), the state space whereby is interpreted as external, i.e., nature chooses an alternative, which has a different implication from ours, where (ex post) self chooses an alternative. The last argument illustrates that the distinction between these two interpretations is subtle, yet very crucial.[14]

9 Concluding Remarks

In this paper, we have discussed the implications of stochastic choice in terms of preference over menus. We have shown that our ASC representation includes many subclasses related to the existing studies, and provided an interpersonal comparison result.

A natural interpretation of stochastic choice is bounded rationality, i.e., the DM makes a suboptimal choice because some psychological effects urge her to deviated from utility maximization. Selten (1975) noted that "[t]here cannot be any mistakes if the players are absolutely rational. Nevertheless, a satisfactory interpretation of equilibrium points ... seems to require that the possibility of mistakes is not completely excluded. This can be achieved by a point of view which looks at complete rationality as a limiting case of incomplete rationality." He also argues that the probability measure over alternatives when the DM possibly makes a mistake is determined by "some unspecified psychological mechanism."

One of our contributions is to provide a decision theoretic foundation to unravel the "unspecified psychological mechanism:" we derive subjective stochastic choice from a preference over menus, rather than exogenously assuming a probability measure over alternatives as is often the case with the existing studies. Thus, we believe that our approach casts a new light on stochastic choice behavior.

[14]Nehring's (1999) argument on the use of convex or concave capacity in comparing a version of his representation with an Choquet expectation and Choquet expected utility (Schmeidler 1989) parallels ours.

Appendix

Proof of Lemma 1

Statement b implying a is straightforward. We show that the converse implication by induction. First, fix an arbitrary singleton menu $x_1 = \{\beta_1\}$. Define $S_1 = \{s_1\}$ and $\phi : \mathcal{A}_1 \times S_1 \to \Delta(Z)$ such that $\phi(x, s_1) = \beta$ for all $x = \{\beta\} \in \mathcal{A}_1$. The state-wise mixture of ϕ_x and ϕ_y with λ and $1 - \lambda$, respectively, clearly equals $\phi_{\lambda x + (1-\lambda)y}$.

Next, suppose that for an arbitrary integer M, there exist S_M and $\phi : \mathcal{A}_M \times S_M \to \Delta(Z)$ that satisfy statement b. Let a menu $x_M = \{\beta_1, \cdots, \beta_M\}$ be such that $\phi(x_M, s_i) = \beta_i$ for $i = 1, \cdots, M$. Then, Assumption 1b implies that there exists a mixture-invariant addition β_{M+1} to x_M. We define $x_{M+1} = x_M \cup \{\beta_{M+1}\}$, $S_{M+1} = \{s_1, \cdots, s_{M+1}\}$, and $\phi(x_{M+1}, s_i) = \beta_i$ for $i = 1, \cdots, M+1$. Conditions a, b, and c of Definition 1 assure that ϕ and S_{M+1} are well-defined. Condition d implies that for all x, $y \in \mathcal{A}_{M+1}$ the state-wise mixture of ϕ_x and ϕ_y with λ and $1 - \lambda$, respectively, equals $\phi_{\lambda x + (1-\lambda)y}$. **Q.E.D.**

Proof of Theorem 1

The sufficiency part is straightforward. We only show the necessity part.

First, note that there exists an affine function u, i.e., $u(\lambda\beta + (1-\lambda)\gamma) = \lambda u(\beta) + (1-\lambda)u(\gamma)$ for all $\lambda \in [0, 1]$, β, $\gamma \in \Delta(Z)$, that represents the commitment ranking $\succsim$.

Lemma 2 There exists an affine function u that represents $\succsim$, i.e., for all β, $\beta' \in \Delta(Z)$, $u(\beta) \geq u(\beta')$ whenever $\{\beta\} \succsim \{\beta'\}$. Further, u is unique up to a positive affine transformation.

Proof The conclusion is straightforward from weak order, continuity, and singleton independence. Note that independence for lotteries follows from singleton independence, and Archimedian continuity clearly follows from our continuity axiom. **Q.E.D.**

The next lemma indicates that there exists a functional J that represents $\succsim$.

Lemma 3 Let u be a affine function derived from Lemma 2. Then, there exists a functional $J : \mathcal{A} \to \Re$ such that (i) for all x, $y \in \mathcal{A}$, $x \succsim y$ if and only if $J(x) \geq J(y)$, and (ii) for all $\beta \in \Delta(x)$ and $x = \{\beta\}$, $J(x) = u(\beta)$.

Proof For singleton menus, J is uniquely defined by (ii). To define J for other menus, fix $\underline{\beta}$ and $\bar{\beta}$ such that $\underline{\beta} < \beta < \bar{\beta}$. (Assumption 1 guarantees that such $\underline{\beta}$ and $\bar{\beta}$ exist.) Then, it follows from continuity that for all $x \in \mathcal{A}_M$, there exists a unique $\lambda \in [0,1]$ such that $x \sim \lambda\underline{\beta} + (1-\lambda)\bar{\beta}$. Thus, by defining $J(x) = J(\lambda\underline{\beta} + (1-\lambda)\bar{\beta})$, J also satisfies (i). **Q.E.D.**

Next, we construct a set of mental states and a choice function. In the meantime, we fix the cardinality of menus M and the function u such that $u(\beta) > 1$ and $u(\beta') < -1$ for some β, $\beta' \in \Delta(Z)$.

Lemma 1 implies that a perfectly correlated mixture coefficient derives the set of mental states S_M and a choice function ϕ. It follows from the construction of ϕ and S_M that there exists a bijection from $\mathcal{A}_M$, the set of menus of cardinality M, to the set of all functions in the form of $\phi_x(\cdot) = \phi(x, \cdot)$ for some $x \in \mathcal{A}_M$, and the corresponding preference order $\succsim'$ such that $x \succsim y$ if and only if $\phi_x \succsim' \phi_y$.

Now, we denote by B a linear space of all functions $a : S_M \to \Re$, endowed with state-wise scholar multiplication and addition. Since S_M is finite, B is equivalent to the linear subspace of simple functions, which we denote by $B_0 \subseteq B$. We also define $K = u(Z)$ and denote by $B_0(K)$ $\subseteq B$ the set of simple functions whose range is K. For $\xi \in \Re$, we denote by $\xi^* \in B$ a constant function such that $\xi^*(s) = \xi$ for all $s \in S_M$.

In the following, we define $\tilde{u} : \mathcal{A} \to K^{S_M}$ by $\tilde{u}(x) = u \circ \phi_x$. It follows from Assumption 1a that the set of such functions $\tilde{u}$ is equivalent to $B_0(K)$. The following lemma characterizes a functional on B_0 that is derived from the axioms. Note that unlike Gilboa and Schmeidler (1989), the functional is sublinear instead of superlinear.

Lemma 4 There exists a functional $I : B_0 \to \Re$ such that
(i) for all $x \in \mathcal{A}_M$, $I(\tilde{u} \circ x) = J(\phi_x)$,
(ii) I is monotonic, i.e., $a \geq b$ implies $I(a) \geq I(b)$ for all a, $b \in B_0$
(iii) I is sublinear (subadditive and homogeneous of degree 1), and
(iv) I is C-independent, i.e., $I(a + \xi^*) = I(a) + I(\xi^*)$ for all $a \in B_0$ and $\xi \in \Re$.

proof First, we define I on $B_0(K)$ by (i). The monotonicity of I (ii) follows from the monotonicity axiom. We indicate that I satisfies (iii)-(iv). Note that (i) implies that $I(1^*) = 1$.

As for superlinearity, we first show $I(\lambda a) = \lambda I(a)$ for all $1 \geq \lambda > 0$ and a, $\lambda a \in B_0(K)$. Let $b = \lambda a$, $y \in \mathcal{A}$ be such that $a = \tilde{u} \circ y$, and $\beta \in \Delta(Z)$ be such that $J(\{\beta\}) = \tilde{u} \circ \{\beta\} = 0$. Then, by defining $x = \lambda y + (1-\lambda)\beta$, $I(\tilde{u} \circ x) = I(\lambda a + (1-\lambda)\tilde{u} \circ \{\beta\}) = I(\lambda a) = I(b)$. Thus, $I(b) = J(x)$. Now, define β' such that $\{\beta'\} \sim y$. Then, by the S-independence axiom, $\lambda\beta' + (1-\lambda)\beta$ $\sim \lambda y + (1-\lambda)\beta = x$. Thus, $J(x) = J(\lambda\beta' + (1-\lambda)\beta) = \lambda J(\beta') + (1-\lambda)J(\beta) = \lambda J(\beta') = \lambda I(\tilde{u} \circ y) = \lambda I(a)$, where the second equality follows from Lemma 3. Accordingly, we obtain $I(b) = I(\lambda a) = \lambda I(a)$.

Now, define $I(a) = \frac{1}{\lambda}I(\lambda a)$ for all $\lambda > 0$ and $\lambda a \in B_0(K)$. Then, by the positive homogeneity on $B_0(K)$, $I(a)$ is homogeneous of degree 1 for all $a \in B$.

Next, we show that I is C-independent. By positive homogeneity that we have shown, it suffices to show that $I(\frac{1}{2}a + \frac{1}{2}\xi^*) = \frac{1}{2}I(a) + \frac{1}{2}I(\xi^*)$ for all $\xi \in \Re$. For $\xi \in \Re$, we define x such that $a = \tilde{u} \circ x$, $\beta' \in \Delta(Z)$ such that $\{\beta'\} \sim x$, and $\beta \in \Delta(Z)$ such that $\tilde{u} \circ \{\beta\} = \xi^*$. Then, by the S-independence axiom, $\frac{1}{2}x + \frac{1}{2}\{\beta\} \sim \frac{1}{2}\{\beta'\}x + \frac{1}{2}\{\beta\}$, which implies that $I(\frac{1}{2}a + \frac{1}{2}\xi^*) = I(\frac{1}{2}\tilde{u} \circ \{\beta'\} + \frac{1}{2}\tilde{u} \circ \{\beta\}) = J(\frac{1}{2}\{\beta'\} + \frac{1}{2}\{\beta\}) = \frac{1}{2}J(\{\beta'\}) + \frac{1}{2}J(\{\beta\}) = \frac{1}{2}I(a) + \frac{1}{2}I(\xi^*)$.

Finally, we show that I is subadditive. By homogeneity, it suffices to show that $I(\frac{1}{2}a + \frac{1}{2}b) \leq \frac{1}{2}I(a) + \frac{1}{2}I(b)$ for a, $b \in B_0(K)$. Let x, $y \in \mathcal{A}_M$ be such that $a = \tilde{u} \circ x$ and $b = \tilde{u} \circ y$. Suppose that $I(a) = I(b)$, i.e., $x \sim y$. Then, it follows from AMI that $x \succsim \frac{1}{2}x + \frac{1}{2}y$, which implies that $I(a) = \frac{1}{2}I(a) + \frac{1}{2}I(b) \geq I(\frac{1}{2}a + \frac{1}{2}b)$. Next, suppose that $I(a) > I(b)$. Define $\xi = I(a) - I(b)$ and $c = b + \xi^*$. Note that $I(c) = I(b + \xi^*) = I(b) + \xi = I(a)$ (the second equality follows from C-independence (iv)). Then, we obtain

$$I(\frac{1}{2}a + \frac{1}{2}b) + \frac{1}{2}\xi = I(\frac{1}{2}a + \frac{1}{2}c) \leq \frac{1}{2}I(a) + \frac{1}{2}I(c) = \frac{1}{2}I(a) + \frac{1}{2}I(b) + \frac{1}{2}\xi,$$

which complete the proof. Note that the first and third equalities follow from C-independence and the second follows from the argument for $I(a) = I(b)$. **Q.E.D.**

The following lemma is the counterpart of Lemma 6.5 by Gilboa and Schmeidler (1989).

Lemma 5 Let I be a monotonic, sublinear, and C-independent functional on B with $I(1^*) = 1$. Then, there exists a closed and convex set $\mathcal{M}_M$ of finitely additive probability measures over S_M such that $I(b) = \max\{\int b d\mu | \mu \in \mathcal{M}_M\}$ for all $b \in B$. Furthermore, $\mathcal{M}_M$ is unique.

Proof Fix $b \in B$ such that $I(b) > 0$. We will first show that there exists a (finitely-additive) probability measure μ_b such that $I(b) = \int b d\mu_b$ and $I(a) \geq \int a d\mu_b$ for all $a \in B$. Define

$$D_1 = \text{conv}(\{a \in B | a > 1^*\} \cup \{a \in B | a \geq b/I(b)\})$$

(where conv(A) denotes the convex hull of the set A) and

$$D_2 = \{a \in B | I(a) < 1\}.$$

Since both D_1 and D_2 are convex, it follows from a separating hyperplane theorem (e.g., Aliprantis and Border 2006) that there exists a linear functional F_b and $\lambda \in \Re$ such that

$$F_b(d_1) \geq \lambda \geq F_b(d_2) \tag{6}$$

for all $d_1 \in D_1$ and $d_2 \in D_2$. Since we clearly have $\lambda > 0$ (otherwise F_b must be identically zero), we set $\lambda = 1$ without loss of generality.

Then, (6) implies that $F_b(1^*) \geq 1$. In addition, since 1^* is a limit point of D_2, the inverse inequality also holds, which concludes $F_b(1^*) = 1$. Further, F_b is nonnegative since for all $E \subseteq S_x$, $E \neq \phi$ and the indicator function 1_E of E, $1^* - 1_E \in D_2$ and $F_b(1_E) + F_b(1^* - 1_E) = F_b(1^*)$ $= 1$.

Accordingly, since F_b is a linear functional, a Riesz representation theorem (Aliprantis and Border 2006) implies that there exists a probability measure μ_b such that $F_b(a) = \int a d\mu_b$ for all $a \in B$. We will show that $F_b(a) \leq I(a)$ for all $a \in B$ and $F_b(a) = I(a)$ for $a = b$. First, assume that $I(a) > 0$. Since $a/I(a) - (1/n)^* \in D_2$ for all $n \in N$ and $F_b(a)$ is continuous with respect to a, we have $F_b(a) \leq I(a)$ from (6). A similar implication for $I(a) \leq 0$ follows from C-independence (set $\xi \in \Re$ so that $I(a + \xi^*) > 0$). Second, we focus on a special case $a = b$. Note that $b/I(b) \in D_1$, which follows from (6) that $F_b(b) \geq I(b)$. Since the previous argument indicates that the inverse inequality also holds, $F_b(b) = I(b)$.

Now, define $\mathcal{M}_M$ by the closure of $\text{conv}\{\mu_b | b \in B, I(b) > 0\}$. Clearly, $I(a) \geq \max\{\int a d\mu | \mu \in \mathcal{M}_x\}$. Since the arguments in the previous paragraphs hold for all $b \in B$, substituting a for b and adapting the definitions of D_1 and D_2 implies for $I(a) > 0$ that there exists a probability measure $\mu_a \in \mathcal{M}_M$ such that $\int a' d\mu_a \geq I(a')$ for all $a' \in B$ and the equality holds for $a' = a$. By applying C-independence, a similar argument also holds for $I(a) \leq 0$.

Now, we show the uniqueness result. Since u is unique up to a positive affine transformation as Lemma 2 indicates, assume $u = u'$ without loss of generality. Condition (v) of Definition 4 (restricted to $\mathcal{A}_M$ and S_M) is clearly satisfied by the construction of ϕ and S_M. Suppose that there exist closed and convex sets $\mathcal{M}_M$ and $\mathcal{M}'_M$ that satisfy the statement in Lemma 5, i.e., $J(\phi_x) = \max_{\mu \in \mathcal{M}_M} \int_{s \in S_M} u(\phi_x(s)) d\mu(s)$ and $J'(\phi_x) = \max_{\mu \in \mathcal{M}'_M} \int_{s \in S_M} u(\phi_x(s)) d\mu(s)$ both represents $\succsim$. Now, choose $\tilde{\mu} \in \mathcal{M} \setminus \mathcal{M}'$. Then, a separating hyperplane theorem indicates that there exists $a \in B$ such that $\int a d\mu > \max_{\mu \in \mathcal{M}'_M} \int_{s \in S_M} a d\mu(s)$. Without loss of generality, we assume that $a \in B_0(K)$. Then, there exists $y \in \mathcal{A}$ such that $J(\phi_y) > J'(\phi_y)$, which contradicts the assumption that J and J' both represent $\succsim$. **Q.E.D.**

Now, we conclude the proof of Theorem 1. Lemmas 2 to 5 imply that for all finite M and $x \in \mathcal{A}_M$, there exists an essentially unique ASC representation. We can define $S = \{s_1, s_2, \cdots\}$ and $\mathcal{M} = \cup_{i=n}^{\infty} \mathcal{M}_n$, by renumbering the elements of each S_M so that $\rho_{x_M, x_N}(\beta_i^M, \beta_i^N) = 1$ and $\phi(x_M, s_i) = \beta_i^M$ for all M, $N = 1, 2, \cdots, 1 \leq i \leq \min\{M, N\}$. Hence, we obtain an ASC representation for all $x \in \mathcal{A}$. **Q.E.D.**

Proof of Theorem 2

The basic idea of the proof is an adaptation of Pires (2002). In the following, we denote by B the set of functions $a : S_n \to \Re$. We define B_0 and $B_0(K)$ in a manner similar to the ones in the proof of Theorem 1.

We first show that Axiom 7 implies $\mathcal{M}^*_{n,k} \subseteq \mathcal{M}_n$. Suppose, to the contrary, that there exists a state-generating function $\mu' \in \Delta(S_n)$ such that $\mu' \in \mathcal{M}^*_{n,k}$ and $\mu' \notin \mathcal{M}_n$. It follows from a separating hyperplane theorem that there exists $a \in B$ such that $\max_{\mu\in\mathcal{M}_n} \int ad\mu < \int ad\mu'$. Without loss of generality, we assume $a \in B_0(K)$. (Note that $B = B_0$ since S_n is finite.) Thus, there exists $\phi : \mathcal{A} \times S_n \to \Delta(Z)$ such that $u \circ \phi(x, \cdot) = a$ and

$$\max_{\mu\in\mathcal{M}_n} \int_{S_n} u(\phi(x,s))d\mu < \int_{S_n} u \circ \phi(x,s)d\mu' = \int_{S_{n+k}} u \circ \phi(x,s)d\mu'. \tag{7}$$

Suppose that $x \sim \{\bar{\beta}\}$, which implies that the left-hand-side of (7) is equal to $u(\bar{\beta})$. Multiplying both sides by λ and adding $(1-\lambda)u(\bar{\beta})$ to them, we obtain

$$u(\bar{\beta}) < \lambda \int_{S_n} u \circ \phi(x,s)d\mu' + (1-\lambda)u(\bar{\beta}) \leq \max_{\mu\in\mathcal{M}_{n+k}} \int_{S_{n+k}} u \circ \phi^{\bar{\beta}}_{n+k}(x,s)d\mu', \tag{8}$$

where $\phi^{\bar{\beta}}_{n+k} : \mathcal{A}\times S_{n+k} \to \Delta(Z)$ is such that $\phi^{\bar{\beta}}_{n+k}(x,s_i) = \phi(x,s_i)$ for $i = 1, \cdots, n$ and $\phi^{\bar{\beta}}_{n+k}(x,s_i)$ $= \bar{\beta}$ for $i = n+1, \cdots, n+k$. However, (8) violates Axiom 7, contradiction.

Next, to prove part (ii), it suffices to show the converse direction for all $\mu \in \mathcal{M}_{n+k}$ such that $\mu(S_n) > 0$ for all n. Suppose, to the contrary, that there exists a state-generating function μ'' $\in \Delta(S_n)$ such that $\mu'' \in \mathcal{M}_n$ and $\mu'' \notin \mathcal{M}^*_{n,k}$. Then, by a separating hyperplane theorem, there exists $a \in B$ such that $\max_{\mu\in\mathcal{M}^*_{n,k}} \int_{S_n} ad\mu < \int_{S_n} ad\mu''$. Without loss of generality, we assume a $\in B_0(K)$. Then, there exists $\phi : \mathcal{A} \times S_n \to \Delta(Z)$ such that

$$\max_{\mu\in\mathcal{M}^*_{n,k}} \int_{S_n} u \circ \phi(x,s)d\mu < \int_{S_n} u \circ \phi(x,s)d\mu''. \tag{9}$$

Suppose that $x \sim \{\bar{\beta}\}$, which implies that

$$u(\bar{\beta}) = \max_{\mu\in\mathcal{M}_n} \int_{S_n} u \circ \phi(x,s)d\mu \geq \int_{S_n} u \circ \phi(x,s)d\mu''.$$

Then, (9) implies that

$$u(\bar{\beta}) > \max_{\mu\in\mathcal{M}^*_{n,k}} \int_{S_n} u \circ \phi(x,s)d\mu = \max_{\mu\in\mathcal{M}_n} \int_{S_n} \frac{u \circ \phi(x,s)}{\mu(S_n)}d\mu. \tag{10}$$

(Note that we assume $\mu(S_n) > 0$ for all $\mu \in \mathcal{M}_{n+k}$.)

The last inequality implies that $u(\bar{\beta}) > \int_{S_n} \frac{u\circ\phi'(x,s)}{\mu(S_n)}d\mu$ for all $\mu \in \mathcal{M}_n$. Accordingly,

$$u(\bar{\beta}) > \mu(S_n) \int_{S_n} \frac{u \circ \phi'(x,s)}{\mu(S_n)}d\mu + (1-\mu(S_n))u(\bar{\beta}),$$

which is equivalent to

$$u(\bar{\beta}) > \max_{\mu \in \mathcal{M}_n} \left[\mu(S_n) \int_{S_n} \frac{u \circ \phi(x,s)}{\mu(S_n)} d\mu + (1 - \mu(S_n)) u(\bar{\beta}) \right] = \max_{\mu' \in \mathcal{M}_{n+k}} \int_{S_{n+k}} u \circ \phi^{\bar{\beta}}_{n+k}(x,s) d\mu',$$

where $\phi^{\bar{\beta}}$ is defined above. The last inequality contradicts Axiom 7. **Q.E.D.**

Proof of Theorem 3

The sufficiency part ($\mathcal{M}_n = \Delta(S_n)$ implies monotonicity) is straightforward. We show the necessity part.

First, we prove the following lemma.

Lemma 6 *Suppose that $\succsim$ has an ASC representation and satisfies monotonicity. If $x \in \mathcal{A}$ is such that $\{\beta\} \subseteq x$ for some $\beta \in \Delta(Z)$ and $\beta \succ \beta'$ for all $\beta' \in x$, then there exists $\mu \in \mathcal{M}_{|x|}$ such that $\mu(s) = 1$ for $s \in S_{|x|}$ with $\phi(x,s) = \beta$.*

Proof Label $\bar{s} \in S_{|x|}$ such that $\phi(x,\bar{s}) = \beta$. Suppose, to the contrary, that for all $\mu \in \mathcal{M}_{|x|}$, $\mu(\bar{s}) < 1$. Since $\succsim$ has an ASC representation, there exists $\mu' \in \mathcal{M}_{|x|}$ such that $V(x) = \int_{S_{|x|}} u(\phi(x,s)) d\mu'$ and $\mu'(\bar{s}) < 1$. Then, $V(x) < u(\beta)$ since $\beta \succ \beta'$ for all $\beta' \in x$, which contradicts monotonicity. **Q.E.D.**

Since Lemma 6 applies for all such β and x, there exists $\mu \in \mathcal{M}_n$ such that $\mu(s) = 1$ for all n and $s \in S_n$. Moreover, since $\mathcal{M}_n$ is closed and convex, $\mathcal{M}_n \supseteq \overline{\text{conv}\{\mu \in \Delta(S_n) | \mu(s) = 1 \text{ for some } s \in S_n\}} = \Delta(S_n)$. It is straightforward that $\mathcal{M}_n \subseteq \Delta(S_n)$, which concludes the proof. **Q.E.D.**

Proof of Theorem 4

The sufficiency part ($\mathcal{M}^1 \supseteq \mathcal{M}^2$ implies that $\succsim_2$ exhibits a preference for commitment more than $\succsim_1$) is straightforward.

We show that $\succsim_2$ prefers commitment more than $\succsim_1$ implies $\mathcal{M}^1 \supseteq \mathcal{M}^2$. Suppose not, i.e., there exists $\mu' \in \mathcal{M}^2$ such that $\mu' \notin \mathcal{M}^1$. Since $\mathcal{M}^1$ is convex, a separating hyperplane theorem implies that there exists $n \in N$ and $a : S_n \to \Re$ such that $\max_{\mu \in \mathcal{M}^1} \int a d\mu < \int a d\mu'$. Without loss of generality, we assume that there exists $\tilde{x} \in \mathcal{A}$ such that $a = u \circ \phi(\tilde{x}, \cdot)$ for the choice function ϕ described in the representation. Then, we have

$$\max_{\mu \in \mathcal{M}^1} \int_{S_{|x|}} u(\phi(\tilde{x},s)) d\mu(s) < \int_{S_{|x|}} u(\phi(\tilde{x},s)) d\mu'(s) \leq \max_{\mu \in \mathcal{M}^2} \int_{S_{|x|}} u(\phi(\tilde{x},s)) d\mu(s).$$

Choosing $\beta \in \Delta(Z)$ such that $\max_{\mu \in \mathcal{M}^1} \int_{S_{|x|}} u(\phi(\tilde{x}, s))d\mu(s) < u(\beta) < \max_{\mu \in \mathcal{M}^2} \int_{S_{|x|}} u(\phi(\tilde{x}, s))d\mu(s)$ leads to contradiction. **Q.E.D.**

Proof of Theorem 5

Note that Axiom 3' implies both S-independence and AMI. First, Axiom 3' implies independence for singleton menus. Thus, as in the proof or Theorem 1, there exists a von Neumann-Morgernstern utility function u that represents the commitment ranking. Further, Assumption 1b implies that for all $M = 1, 2, \cdots, x, y \in \mathcal{A}_M$, and $\tilde{\beta} \in \Delta(Z)$, there exists $z \in \mathcal{A}_M$ such that $\beta = \tilde{\beta}$ for all $\beta \in z$, which implies S-independence together with Axiom 3'. The implication of AMI is straightforward.

Thus, from the proof of Theorem 1, there exists a bijection from $\mathcal{A}_M$ to the set of all choice functions $\phi_x : S_{|x|} \to x$ for $x \in \mathcal{A}_M$. Hence, the independence axiom on $\succsim$ implies the independence axiom on the corresponding preference $\succsim'$ over choice functions (with respect to state-wise convex combination), the conclusion follows from Anscombe and Aumann (1963). **Q.E.D.**

Proof of Theorem 6

The sufficiency part is straightforward.

Conversely, it follows from Theorem 4 that $\succsim$ has a pure stochastic choice representation in which $\mathcal{M}_n = \{\mu_n\}$ for some $\mu_n \in \Delta(S_n)$, $n = 1, 2, \cdots$. Axiom 8 implies that $\mu_n(s_i) \geq \mu_n(s_j)$ for all $i, j, n \in N$. **Q.E.D.**

Proof of Theorem 7

The sufficiency part is straightforward.

Conversely, suppose that there exist $n \in N$ and $\mu \in \Delta(S_n)$ such that $\mu(s_i) > 0$ for $i \neq 1$. Now, let $\beta \in \Delta(Z)$ and $x = \{\beta_1, \cdots, \beta_n\}$ such that $\beta_1 \succ^* \cdots \succ^* \beta_n$, β_i is of the highest commitment ranking in x, and $\beta_i \succ \beta \succ \beta_1$. Then, by the definition of an ASC representation, we have $x \succ \{\beta\}$, yet $\beta \succ \beta_1$, contradiction. **Q.E.D.**

References

Aliprantis, C.D. and Border, K.C. (2006), *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third edition, Springer Verlag, Berlin.

Anscombe, F.J. and Aumann, R. (1963), A Definition of Subjective Probability, *Annals of Mathematical Statistics* 34, 199-205.

Bandyopadhyay, T., Dasgupta, I., Pattanaik, P.K. (1999), Stochastic Revealed Preference and the Theory of Demand, *Journal of Economic Theory* 84, 95-110.

Broadbent, D.E., Cooper, P.F., FitzGerald, P., and Parkes, K.R. (1982), The Cognitive Failures Questionnaire (CFQ) and Its Correlates, *British Journal of Clinical Psychology* 19, 177-188.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2010), Salience Theory of Choice under Risk, *mimeo.*

Chatterjee, K. and Krishna, R.V. (2009), *A "Dual Self" Representation for Stochastic Temptation,* American Economic Journal: Microeconomics 1, 148-167.

Davidson, D. and Marschak, J. (1959), Experimental Tests of a Stochastic Decision Theory, in C.W. Churchman (ed.), *Measurement Definitions and Theories.* New York, NY: John Wiley & Sons, pp. 233-269.

Debreu, G. (1958), Stochastic Choice and Cardinal Utility, *Econometrica* 26, 440-444.

Dekel, E., Lipman, B., and Rustichini, A. (2001), Representing Preferences with a Unique Subjective State Space, *Econometrica* 69, 891-934.

Dekel, E., Lipman, B., and Rustichini, A. (2009), Temptation-Driven Preferences, *Review of Economic Studies* 76, 937-971.

Dekel, E. and Lipman, B. (2010), Costly Self Control and Random Self Indulgence, *mimeo.*

Epstein, L., Marinacci, M., and Seo, K. (2007), Coarse Contingencies and Ambiguity, *Theoretical Economics* 2, 355-394.

Ergin, H. and Sarver, T. (2010), A Unique Costly Contemplation Representation, *Econometrica*, forthcoming.

Fudenberg, D. and Levine, D. (2006), A Dual Self Model of Impulse Control, *American Economic Review* 96, 1449-1476.

Ghirardato, P. and Marinacci, M. (2002), Ambiguity Made Precise: A Comparative Foundation, *Journal of Economic Theory* 102, 251-289.

Gilboa, I. and Schmeidler, D. (1989), Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics* 18, 141-53.

Gul, F. and Pesendorfer, W. (2001), Temptation and Self-Control, *Econometrica* 69, 1403-1435.

Gul, F. and Pesendorfer, W. (2006), Random Expected Utility, *Econometrica* 74, 121-146.

Harless, D. and Camerer, C. (1994), The predictive utility of generalized utility theories, *Econometrica* 62, 1251-89.

Huber, J., Payne, J., and Puto, C. (1982), Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis, *Journal of Consumer Research* 9, 90-98.

Kahneman, D. (1994), New Challenges to the Rationality Assumption, *Journal of Institutional and Theoretical Economics* 150, 18-36.

Kahneman, D., Wakker, P.P., and Sarin, R. (1997), Back to Bentham? Explorations of Experienced Utility, *Quarterly Journal of Economics* 112, 375-405.

Koopmans, T.C. (1964), On the flexibility of future preferences, in Shelly, M.W. and Bryan, G.L. (eds.) *Human judgments and optimality*, John Wiley and Sons, New York.

Kreps, D.M. (1979), A Representation Theorem for "Preference for Flexibility", *Econometrica* 47, 565-577.

Kreps, D.M. (1992), Static Choice and Unforeseen Contingencies, in Dasgupta, P., Gale, D., Hart, O., and Maskin, E. (eds.) *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn*, MIT Press, Cambridge, MA, 259-281.

Loewenstein, G. (1996), Out of Control: Visceral Influences on Behavior, *Organizational Behavior and Human Decision* 65, 272-292.

Loewenstein, G. and O'Donoghue (2004), Animal Spirits: Affective and Deliberative Processes in Economic Behavior, *mimeo.*

Luce, R.D. (1959), *Individual choice behavior*, John Wiley & Sons Inc., New York.

Luce, R.D. and Suppes (1965), Preference, Utility and Subjective Probability, in Luce, R.D., Bush, R.R., Galanter, E. (eds.) *Handbook of Mathematical Psychology* Vol III, Wiley, New York.

Manzini, P. and Mariotti, M. (2010), A Salience Theory of Choice Errors, *mimeo.*

Masatlioglu, Y., Nakajima, D., and Ozbay, E.Y. (2010), Revealed Attention, *mimeo.*

McFadden, D.L. and Richter, K. (1991), Stochastic Rationality and Revealed Stochastic Preference, in J. Chipman, D. McFadden, K. Richter (eds) *Preferences, Uncertainty, and Rationality*, Westview Press, 161-186.

Nehring, K. (1999), Preference for Flexibility in a Savage Framework, *Econometrica* 67, 101-119.

Nishimura, K.G. and Ozaki, H. (2006), An Axiomatic Approach to Epsilon-contamination, *Economic Theory* 27, 333-340.

Noor, J. and Takeoka, N. (2009), Menu-Dependent Self-Control, *mimeo.*

Noor, J. and Takeoka, N. (2010), Uphill Self-Control, *mimeo.*

Olszewski, W. (2007), Preferences over Sets of Lotteries, *Review of Economic Studies* 74, 567-

595.

Pires, C.P. (2002), A rule for updating ambiguous beliefs, *Theory and Decision* 53, 137-152.

Sarver, T. (2008), Anticipating Regret: Why Fewer Options May be Better, *Econometrica* 76, 263-305.

Selten, R. (1975), Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games, *International Journal of Game Theory* 4, 25-55.

Stovall, J.E. (2010), Multiple Temptations, *Econometrica* 78, 349-376.

Strotz, R.H. (1955), Myopia and Inconsistency in Dynamic Utility Maximization, *Review of Economic Studies* 23, 165-180.

Thaler, Richard H. and Shefrin, H.M. (1981), An Economic Theory of Self-Control, *Journal of Political Economy* 89, 392-406.

Wright, P. and Barbour, F. (1977), Phased Decision Strategies: Sequels to an Initial Screening, in Starr, M. and Zeleny, M. (eds.), *Studies in Management Sciences, Multiple Criteria Decision Making*, North-Holland, Amsterdam, pp. 91-109.