Noname manuscript No.
(will be inserted by the editor)

Anticipated stochastic choice

Nobuo Koida

Version: December 9, 2016

Abstract The purpose of this study is to characterize stochastic choice influenced by the objective or subjective *positions* of alternatives in a menu. The main theorem axiomatizes the *anticipated stochastic choice (ASC)* representation, wherein the decision maker maximizes the expected utility by the *cognitive control* of a probability measure over *mental states* that trigger the ex post choice of alternatives. A key prerequisite for this axiomatization is that the randomization between menus is identified with their *perfectly correlated mixture*, which includes only mixtures of specific alternative pairs. The essential uniqueness of an ASC representation defines an index of rationality that is relevant to a preference for commitment. Special cases of ASC include exact utility maximization, uncontrolled stochastic choice, trembling hands, and choice with limited attention. Furthermore, ASC accommodates potentially stochastic choice anomalies such as the attraction effect, cyclical choice, and position effects.

Keywords: menu preference · position-dependence · stochastic choice · cognitive control · preference for commitment · limited attention

JEL Classification Numbers: D81

I am grateful to Itzhak Gilboa, Peter Hammond, Youichiro Higashi, Kazuya Hyogo, Atsushi Kajii, Paola Manzini, Marco Mariotti, Hitoshi Matsushima, Daisuke Nakajima, Yutaka Nakamura, Christopher Olivola, Daisuke Oyama, Hiroyuki Ozaki, Marciano Siniscalchi, Norio Takeoka, and the conference and seminar participants at SAET 2015, JEA 2011 Fall Meeting, Kyoto University, the University of Tokyo, Keio University, Okayama University, and Tohoku University for their helpful discussions and comments. I also thank two anonymous referees for their valuable comments that have significantly improved the paper. Financial support from the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (Grant Numbers 24730168, 26380241) is gratefully appreciated. The usual disclaimer applies.

Faculty of Policy Studies, Iwate Prefectural University, 152-52 Sugo, Takizawa, Iwate 020-0693, Japan. Tel: +81-19-694-2814, Fax: +81-19-694-2701, E-mail: nobuo@iwate-pu.ac.jp

1 Introduction

People often make mistakes. In contrast to the basic assumption of traditional economics that a decision maker (DM) is rational, numerous studies have indicated that DMs make suboptimal choices with a positive probability, owing to various psychological or cognitive effects (Selten 1975).

Many mistakes are *position-dependent*; that is, the choice of one alternative from the set of alternatives, or *menu*, may depend on the order of the alternatives in the menu.[1] For example, various psychological effects suggest that the alternatives presented in a certain position, such as those at the beginning, middle, and end of the menu, are chosen more frequently than others (e.g., Bruine de Bruin 2005; Christenfeld 1995; Murdock 1962). It has also been reported that adding, removing, and replacing items in the menu may draw the DM's attention to alternatives that are superior in a specific attribute (e.g., the quality, salience, or justifiability of the product) rather than their overall utility, which significantly affects the choice of alternative (Huber et al. 1982; Simonson 1989; Slovic 1975; Tyszka 1983). Another example is choice with limited attention (Manzini and Mariotti 2014, 2015; Masatlioglu et al. 2012; Wright and Barbour 1977), which typically restricts the DM's attention to alternatives in a certain part of the menu, such as the top n according to some criteria. Finally, such positions may not only be determined by an objective order (e.g., location, time, or size), but also by a subjective order (e.g., attention, memory, or familiarity) or an outside agent, such as a consultant or search engine, as argued by Rubinstein and Salant (2006).

Under the influence of these mistakes, the DM may exert *cognitive control*, that is, the control of the *probability measure* over mental states that affect choice, rather than directly controlling the choice of alternative (Posner and Snyder 1975). For example, suppose that, in a restaurant serving menu $x = \{beef, fish\}$, the choice of alternative is driven by the mental state: mental state s_1 triggers the choice of the unfavorable $beef$, whereas mental state s_2 triggers the choice of the preferred $fish$. Because the realization of mental state is involuntary, the DM may not be able to fully control the ex post choice in anticipation of this; however, she may still be able to perform some cognitive activities, such as staying focused on the decision at hand and suppressing an unwanted psychological response, to maximize the probability of mental state s_2 being realized, which is associated with the favorable *fish*. A similar idea has been documented under the names of *mood regulation* (Larsen 2000), *emotion regulation* (Gross 1998), and *self-control of automatic associations and behavioral impulses* (Sherman et al. 2008).

The present study formalizes this type of behavior in a menu preference framework. Our main theorem (Theorem 1) axiomatizes the *anticipated stochastic choice (ASC) representation*, which assumes that the DM selects the opti-

[1] We use the term "position-dependence" rather than the more general term "menu-dependence" because the former provides a natural interpretation for the correlated choices of alternatives and potentially full support stochastic choice function, which are key characterizations of the following analysis.

mal probability measure over mental states from a certain set $\mathcal{M}$, to maximize the expected utility.

This model has several distinctive properties. First, the position-dependence of ASC motivates the axiom of *perfectly correlated mixtures of menus*; that is, the DM identifies a *random* menu (i.e., lottery over menus) with a menu that consists of the mixtures of *specific* alternative pairs: for menus x and y with cardinality 2, suppose that mental states s_1 and s_2 are position-relevant; that is, s_1 and s_2 respectively trigger the choices of the first and the second alternatives from each menu. We also assume that the mental states are realized *before* a randomization between x and y is performed, which reflects the idea of automatic choice. Then, the first alternative is chosen from both x and y if mental state s_1 is realized, whereas the second alternative is chosen from both x and y if mental state s_2 is realized. Accordingly, the DM would expect her choices of alternatives from each menu to be perfectly correlated, which is an instrument to derive potentially full support probability measures over mental states. In particular, under our model, a randomization between identical menus x is identified with the original nonrandom menu x. This is in stark contrast to the axiom used in existing studies (e.g., Dekel et al. 2001; Gul and Pesendorfer 2001), which identifies a randomization between the identical menus x with their *uncorrelated mixture* (i.e., Minkowski sum), as we will discuss later.

Second, because the choice of alternative in ASC depends on the menu, the induced stochastic choice over alternatives generally violates standard properties such as regularity (Luce and Suppes 1965) and/or the weak axiom of revealed stochastic preference (WARSP) (Bandyopadhyay et al. 1999). This property also enables us to accommodate various choice anomalies, such as the attraction effect, cyclical choice, and other position effects, as demonstrated in Section 6.1.

Third, Theorems 2 and 3 indicate that ASC exhibits a preference for *commitment to a singleton menu*, and has no nontrivial intersections with the models of a preference for flexibility (Dekel et al. 2001; Kreps 1979; Nehring 1999), despite their similarity. That is, in anticipation of her own mistakes, the consumer in the restaurant example would ex ante prefer menu $y = \{fish\}$ to menu $x = \{beef, fish\}$, to preclude the possibility of erroneously choosing the nonpreferred *beef*.[2] Accordingly, our model portrays position-dependent stochastic choice, rather than temptation (Chatterjee and Krishna 2009; Dekel and Lipman 2012; Dekel et al. 2009; Gul and Pesendorfer 2001; Noor and Takeoka 2010, 2015) and ex post regret (Sarver 2008), as a potential source of preference for commitment.

[2] Anticipated mistakes (i.e., ex ante awareness of ex post mistakes) have been discussed both in psychology (Broadbent et al. 1982) and in economics (Kahneman et al. 1997; Manzini and Mariotti 2015; Piccione and Rubinstein 1997). Gross (1998) also argued that under the influence of emotion, the DM often applies the *situation selection* strategy, that is, she attempts to avoid a situation that possibly creates a negative psychological state leading to an undesirable future choice, which can be interpreted as a preference for commitment.

Finally, the uniqueness of the set $\mathcal{M}$ of probability measure over mental states allows for an interpersonal comparison. Restricting $\mathcal{M}$ yields a wide range of special cases such as exact utility maximization, uncontrolled stochastic choice (i.e., stochastic choice generated by a single probability measure), trembling hands, and choice with limited attention. Theorem 3 also indicates that the size of $\mathcal{M}$ is relevant to a comparative preference for commitment, and provides a behavioral foundation for Selten's (1975) classic argument that regards the error rate in the trembling-hand model as an index of rationality. This also conforms with the argument that higher cognitive abilities reduce vulnerability to psychological effects such as the attraction and position effects (Krueger and Salthouse 2011; Sherman et al. 2008; Tentori et al. 2001).

The remainder of this study is organized as follows: Section 2 describes the basic model that forms the focus of this study. Section 3 presents the basic axioms, and Section 4 states the main representation theorem. Section 5 explores the implications of ASC with respect to size-related properties such as monotonicity. Section 6 discusses our model, and Section 7 concludes the paper.

2 Model

2.1 Preliminaries

Let Z be a finite set of prizes. $\Delta(Z)$ denotes the set of alternatives, or probability distributions over prizes ($\Delta(\cdot)$ denotes the set of finite probability distributions over $(\cdot)$). Let $\mathcal{A} = \mathcal{K}_0(\Delta(Z))$ be the set of finite menus ($\mathcal{K}_0(\cdot)$ denotes the set of all finite subsets of $(\cdot)$) and $\mathcal{A}_n \subseteq \mathcal{A}$ be the set of menus with cardinalities less than $n \in \mathbb{N}$, that is, $\mathcal{A}_n = \{x \in \mathcal{A} : |x| \leq n\}$. We refer to $\Delta(\mathcal{A})$ $= \Delta(\mathcal{K}_0(\Delta(Z)))$ as the set of (finite) random menus. The generic elements in $\Delta(\mathcal{A})$, $\mathcal{A}$, and $\Delta(Z)$ are denoted by P, Q, R, $\cdots$, x, y, z, $\cdots$, and α, β, γ, $\cdots$, respectively.[3] We use $\lambda P \oplus (1-\lambda)Q$ to denote the randomization between random menus P and Q with probabilities λ and $1-\lambda$, respectively.[4] We also define $\bar{n}(P)$, the maximum cardinality of the menus in the support of random menu P; that is, $\bar{n}(P) = \max_{x \in \mathrm{supp}(P)} |x|$ ($\mathrm{supp}(\cdot)$ denotes the support of probability measure $(\cdot)$). Note that $\bar{n}(P)$ is well-defined because P is finite.

This study assumes the preference relation $\succsim$ over $\Delta(\mathcal{A})$, the set of all finite random menus.[5] Clearly, $\succsim$ has a restriction on $\mathcal{A}$. Furthermore, we refer to

[3] As usual, menu x is naturally identified with a degenerated random menu that generates x with a probability of one, and an alternative β is identified with a singleton menu $\{\beta\}$.

[4] By definition, we have $\lambda P \oplus (1-\lambda)Q = (1-\lambda)Q \oplus \lambda P$.

[5] We restrict our attention to finite (random) menus because of a key characterization of the following analysis, namely, a surjection or bijection from one menu to the other can be consistently defined for all menu pairs; this is not generally the case for (uncountably) infinite menus. Furthermore, many studies have examined a preference over (the convex hulls of) finite menus (e.g., Chatterjee and Krishna 2009; Dekel et al. 2009; Gul and Pesendorfer 2001).

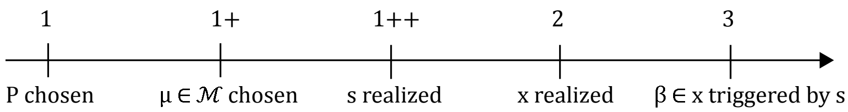

Fig. 1 Timeline

the restriction of $\succsim$ on $\Delta(Z)$ as the *commitment ranking*. For simplicity, we denote $\beta \succsim \gamma$ for all β, $\gamma \in \Delta(Z)$, instead of $\{\beta\} \succsim \{\gamma\}$.

2.2 Timeline

Here, we explain the timeline assumed in this study.[6] The DM chooses a random menu P in period 1, and P generates a (finite) menu x in period 2. The choice of alternative $\beta \in x$ in period 3 is triggered by mental state s, which is realized between periods 1 and 2. In anticipation of this, the DM chooses the optimal probability measure over mental states, μ, from a certain set $\mathcal{M}$ in period 1+, which generates mental state s in period 1++ (Fig. 1).

Our approach differs significantly from those in existing studies (e.g., Dekel et al. 2001) in its timing of the realization of mental states and utility maximization. Existing studies highlight the utility-maximizing choice of an alternative given an ex post preference change, and thus assume that a subjective state is realized *after* a menu has been generated by a random menu, *after* which the state-dependent utility is maximized. In contrast, our primary focus is on the DM's ex ante control of the probability measure over mental states, whereas the realization of a mental state triggers an automatic, rather than deliberative, ex post choice of alternative. Thus, we assume that a mental state s is realized *before* menu x has been generated by random menu P, *before* which a probability measure μ over mental states is chosen to maximize the expected utility. This difference in timing explains our adoption of the terminology "mental states" instead of "subjective states."

The above timeline provides a general framework for discussing various cases, including the following examples. First, psychological effects such as primacy and recency may trigger the choice of an alternative in a specific position on the menu. In this case, mental states are presumably realized before the menu is generated by a random menu, because these psychological effects are involuntary and irrevocable. Second, choice with limited attention (Manzini and Mariotti 2014, 2015; Masatlioglu et al. 2012; Wright and Barbour 1977) may confine the DM's attention to an alternative in a specific position (e.g., the first item) on the menu. Our timeline is also generated in this case, because the alternative in the relevant position is eventually chosen, irrespective of the menu generated by a random menu. Finally, before a menu has been generated by a random menu, temptation may trigger the choice of alternative in a specific position, or conversely, the DM may decide to use positions in the

[6] The timeline outlined here is an interpretation, rather than part of the model. The only item that is assumed to be observable is choice over random menus in period 1.

menu as a commitment device to avoid temptation. In both cases, the choice of the alternative in a specific position is predetermined once a mental state has been realized, which also generates the timeline described above.

3 Axioms

This section states the axioms imposed on the preference relation $\succsim$. The first is standard.

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

The next axiom describes the DM's perceptions of random menus under position-dependent choice, which distinguishes our approach from those used in existing studies. This axiom is characterized by the following steps.

First, in the restaurant example of the Introduction, suppose that the DM regards alternatives *beef* and *fish* in menu $x = \{beef, fish\}$ as being positioned first and second under some criteria (e.g., the alternatives are physically listed or the DM recognizes the alternatives in this order), and thus position-relevant mental states s_1 and s_2 trigger the choices of *beef* and *fish*, respectively. Now, consider a random menu $P = \lambda x \oplus (1-\lambda)x$ for some $\lambda \in [0, 1]$, which generates an identical menu x both in the "λ" and "$1 - \lambda$" events. Because the timeline assumes that the mental states are realized before a menu has been generated by random menu P, mental state s_1 triggers the choice of *beef* in the "λ" event *if and only if* it triggers the same choice in the "$1-\lambda$" event, and similarly for mental state s_2 and *fish*. Thus, the DM expects that her choices of alternatives from each menu x will be *perfectly correlated*, and thus, she identifies the random menu P with the menu that comprises $\lambda beef + (1-\lambda)beef = beef$ and $\lambda fish + (1-\lambda) fish = fish$, namely, the original nonrandom menu x. Because a similar discussion holds for an arbitrary menu, we have $\lambda x \oplus (1 - \lambda)x \sim x$ for all $x \in \mathcal{A}$ and $\lambda \in [0, 1]$.

Next, to extend this inference to randomizations between nonidentical menus, we impose the following two consistency conditions. As we will indicate, these are also satisfied in existing studies such as Dekel et al. (2001) and Gul and Pesendorfer (2001). The first condition requires that the DM form the correct expectation of her choices from each menu generated by a random menu. Consider a random menu $P = \lambda x \oplus (1 - \lambda)y$ over some menus x and y. If she correctly anticipates that she will choose alternative pair (α, β) from menu pair (x, y), she would also expect the mixture $\lambda\alpha + (1 - \lambda)\beta$ of the alternatives to be obtained from P. Because a similar argument holds for other alternative pairs that can be chosen simultaneously, there should exist a nonempty $C \subseteq x \times y$ such that $P \sim \{\lambda\alpha + (1-\lambda)\beta : (\alpha, \beta) \in C\}$. Second, a DM who has consistent perceptions of random menus would expect the choices of alternatives to be correlated *across* different menu pairs. That is, for all menus x, y, $z \in \mathcal{A}$, if the DM expects alternative pairs (α, β) and (β, γ) to be chosen from menu pairs (x, y) and (y, z), respectively, she would also expect alternative pair (α, γ) to be chosen from menu pair (x, z), because there must exist a mental state that triggers the choices of alternatives α, β, and γ from menus x, y, and z, respectively. These two conditions can be formalized as follows.

Definition 1 Preference $\succsim$ *has consistent perceptions of random menus* if the following conditions are satisfied.

(a) For all $x, y \in \mathcal{A}$, there exists a nonempty set $C \subseteq x \times y$ such that $\lambda x \oplus (1-\lambda)y \sim \lambda x +_C (1-\lambda)y \equiv \{\lambda\alpha + (1-\lambda)\beta : (\alpha, \beta) \in C\}$ for all $\lambda \in [0, 1]$.
(b) For all x, y, $z \in \mathcal{A}$ such that $|x| \leq |y| \leq |z|$, $C \subseteq x \times y$, and $C' \subseteq y \times z$, if $\lambda x \oplus (1-\lambda)y \sim \lambda x +_C (1-\lambda)y$ and $\lambda' y \oplus (1-\lambda')z \sim \lambda' y +_{C'} (1-\lambda')z$ for all λ, $\lambda' \in [0, 1]$, then $\lambda'' x \oplus (1-\lambda'')z \sim \lambda'' x +_{C''} (1-\lambda'')z$ for all $\lambda'' \in [0, 1]$, where $C'' = \{(\alpha, \gamma) : (\alpha, \beta) \in C, (\beta, \gamma) \in C'\}$.

The restriction $|x| \leq |y| \leq |z|$ in condition (b) is imposed to obtain a class of menu mixtures that are more general than the Minkowski sum; if we drop this restriction and consider a singleton menu $y = \{\hat{\beta}\}$, for example, we may have $(\alpha, \hat{\beta}) \in C$ and $(\hat{\beta}, \gamma) \in C'$ for all x, $z \in \mathcal{A}$, $\alpha \in x$, and $\gamma \in z$ in condition (b). Then, the condition implies that a randomization between x and z is trivially indifferent to their Minkowski sum, that is, $C'' = \{(\alpha, \gamma) : \alpha \in x, \gamma \in z\}$.

The following axiom summarizes this discussion:

Axiom 2 (Perfectly correlated mixtures of menus)
(a) For all $x \in \mathcal{A}$ and $\lambda \in [0, 1]$, $\lambda x \oplus (1-\lambda)x \sim x$.
(b) $\succsim$ has consistent perceptions of random menus.

A typical situation wherein Axiom 2 holds is that in which each alternative in the menu is explicitly indexed (e.g., by its physical location) and correlation occurs among alternatives with specific indices, typically, the i-th alternatives in each menu. Formally, for all menus $x = \{\alpha_1, \cdots, \alpha_m\}$ and $y = \{\beta_1, \cdots, \beta_n\}$ with $m \leq n$, letting $C = \{(\alpha_{\min\{i,m\}}, \beta_i) : i = 1, \cdots, n\}$ and $\lambda x \oplus (1-\lambda)y \sim \lambda x +_C (1-\lambda)y$ for all $\lambda \in [0, 1]$ satisfies the axiom. Note that, in contrast to Rubinstein and Salant (2006), who considered the set of *lists* (i.e., choice sets with explicitly indexed alternatives) as the domain of choice, our model does not assume such indices of alternatives as a preliminary and allows for correlated choice with respect to subjective (i.e., unobservable) indices determined by, for example, attention, memory, and familiarity.

Unlike our approach, existing literature on menu preferences (e.g., Dekel et al. 2001; Gul and Pesendorfer 2001) implicitly imposes the following axiom, which identifies a randomization between menus with the Minkowski sum (see Dekel et al.'s discussion on the independence axiom for details).

Axiom 2' (Uncorrelated mixtures of menus) For all x, $y \in \mathcal{A}$ and $\lambda \in [0, 1]$, $\lambda x \oplus (1-\lambda)y \sim \lambda x + (1-\lambda)y \equiv \{\lambda\alpha + (1-\lambda)\beta : \alpha \in x, \beta \in y\}$.

This axiom is in stark contrast to Axiom 2, particularly when $x = y$; that is, it implies $\lambda x \oplus (1-\lambda)x \sim \lambda x + (1-\lambda)x = \{\lambda\alpha + (1-\lambda)\beta : \alpha, \beta \in x\} \neq x$ for all $\lambda \in (0,1)$. In other words, a randomization between identical menus x is identified with the menu comprising mixtures of *any* two alternatives from x, because the choice of alternative in the "λ" event is uncorrelated with that in the "$1-\lambda$" event.

Section 6.2 contains a further discussion on the implications of Axioms 2 and 2'. Particularly, Axiom 2' has another equivalent representation, Axiom 2", which replaces Axiom 2a with its counterpart while retaining Axiom 2b. In other words, Axiom 2b produces a weak consistency condition that is satisfied not only by Axiom 2, but also by Axiom 2'.[7]

The next continuity axiom is standard, but is restricted here to menus with identical cardinality.

Axiom 3 (Archimedean continuity for menus with constant cardinalities) For all x, y, $z \in \mathcal{A}$ such that $|x| = |y| = |z|$ and $x \succ y \succ z$, there exist λ, $\lambda' \in (0,1)$ such that $\lambda x \oplus (1-\lambda)z \succ y \succ \lambda' x \oplus (1-\lambda')z$.

We permit some discontinuity in the preference for randomizations between menus with different cardinalities, as this allows us to capture possible changes in the menu evaluation due to a change of the perceived positions of alternatives in each menu.

The next axiom is the independence axiom restricted to randomizations with singleton menus.

Axiom 4 (Singleton independence (S-independence)) For all x, $y \in \mathcal{A}$ and $\beta \in \Delta(Z)$, $x \succsim y$ if and only if $\lambda x \oplus (1-\lambda)\{\beta\} \succsim \lambda y \oplus (1-\lambda)\{\beta\}$ for all $\lambda \in [0,1]$.

The intuition behind this axiom is given below. Let x, y, and z be menus and $\lambda \in [0,1]$. The preference ranking between $\lambda x \oplus (1-\lambda)z$ and $\lambda y \oplus (1-\lambda)z$ only differs from the ranking between x and y if randomizing with z affects the evaluations of x and y differently. However, if z is a singleton menu, that is, $z = \{\beta\}$ for some $\beta \in \Delta(Z)$, then all alternatives in x and y will be randomized with the identical alternative β, which uniformly changes the evaluation of alternatives. Thus, the ranking between x and y and that between $\lambda x \oplus (1-\lambda)\{\beta\}$ and $\lambda y \oplus (1-\lambda)\{\beta\}$ should be consistent.

This axiom is similar to those of C-independence (Gilboa and Schmeidler 1989), set S-independence (Olszewski 2007), and independence of degenerated

[7] Whether we can characterize an informative preference representation with *imperfectly* correlated mixtures of menus (i.e., a randomization between identical menus not being indifferent to the menu itself or the Minkowski sum) is an open question. However, this problem may be solved by imposing certain axioms. For example, the set betweenness proposed by Gul and Pesendorfer (2001) renders the preference representation dependent only on two alternatives in a menu, implying that a randomization between menus is indifferent to a doubleton set comprising the mixtures of two specific alternative pairs. Similar axioms may characterize more general cases.

decisions (Ergin and Sarver 2010). In particular, the last axiom can be adapted to our model as $\lambda x \oplus (1-\lambda)\{\beta\} \succsim \lambda y \oplus (1-\lambda)\{\beta\}$ whenever $\lambda x \oplus (1-\lambda)\{\gamma\} \succsim \lambda y \oplus (1-\lambda)\{\gamma\}$ for all x, $y \in \mathcal{A}$, $\beta, \gamma \in \Delta(Z)$, and $\lambda \in [0,1]$, which is a weakening of S-independence. The imposition of the weaker axiom derives a more general representation, wherein choosing the optimal probability measure over mental states involves positive costs. However, this study imposes the S-independence axiom instead of the weaker axiom because it renders our model comparable with various stochastic choice models. In particular, ASC can naturally relate stochastic choices for menus with different cardinalities by applying Bayes' rule (see Online Resource 1 for details), which cannot be obtained under the weaker axiom.

The next axiom requires the DM to prefer menu x to a randomization between x and an inferior menu.

Axiom 5 (Aversion to randomizations with inferior menus (ARI)) For all x, $y \in \mathcal{A}$ such that $|x| = |y|$, if $x \succsim y$, then $x \succsim \lambda x \oplus (1-\lambda)y$ for all $\lambda \in [0,1]$.

To understand the reasoning behind this axiom, let menus x and y be such that $x \succsim y$, and alternatives $\hat{\alpha}$ and $\hat{\beta}$ be of the highest commitment ranking in x and y, respectively. If the DM performs exact utility maximization, we clearly have $x \succsim \lambda x \oplus (1-\lambda)y$, because $x \succsim y$ implies $\hat{\alpha} \succsim \hat{\beta}$ and the alternative chosen from $\lambda x \oplus (1-\lambda)y$ will be at best as good as $\lambda\hat{\alpha} + (1-\lambda)\hat{\beta}$, which is worse than $\hat{\alpha}$. On the other hand, if suboptimal alternatives are chosen from a menu with a positive probability, we may have $x \succ \lambda x \oplus (1-\lambda)y$ even if $\hat{\alpha} \sim \hat{\beta}$, because $x \succsim y$ implies that the suboptimal alternatives in y worsen the evaluation of the menu more than those in x. Thus, randomizing with y may lower x's evaluation, which motivates the axiom.

ARI may appear similar to Gilboa and Schmeidler's (1989) uncertainty aversion and Ergin and Sarver's (2010) aversion to contingent planning (ACP). In particular, the latter axiom characterizes a menu preference as ARI. However, ACP and ARI have different behavioral implications: ACP argues that arranging a choice plan for the mixture of two menus is more cognitively costly than for the individual menus, because it assumes *uncorrelated* mixtures of the menus. That is, the mixture of menus includes the mixtures of all possible alternative pairs, which drastically increases the number of contingencies to be considered before arranging a choice plan for the mixture of menus. In contrast, our model identifies a randomization between menus with their *perfectly correlated* mixture of menus, which only includes the mixtures of correlated alternative pairs. Moreover, the correlated choice in our model presumably involves no cognitive cost, because it is induced by an automatic, rather than deliberate, process driven by mental states. Accordingly, arranging a choice plan for the mixture of menus will be no more cognitively costly than for the individual menus in our model, and thus, ARI is more relevant to the DM's attitude toward randomizations than to cognitive costs.

We restrict ARI to menus with a common cardinality for a similar reason to that given for Axiom 3. For example, for some x, $y \in \mathcal{A}$ such that $|x| < |y|$ and $x \succsim y$, we allow $\lambda x \oplus (1-\lambda)y$ to be strictly preferred to x, because randomizing with menu y of a different cardinality may improve x's evaluation by affecting the positions of alternatives in each menu.

Finally, the following two axioms are technical. First, we state the concept of dominance with respect to correlated alternatives.

Axiom 6 (Dominance) For all x, $y \in \mathcal{A}$, assume that $C \subseteq x \times y$ is such that $\lambda x \oplus (1-\lambda)y \sim \lambda x +_C (1-\lambda)y$ for all $\lambda \in [0,1]$. If $\alpha \succsim \beta$ for all $(\alpha, \beta) \in C$, then $x \succsim y$.

This axiom implies that menu x is preferred to menu y if $\alpha \in x$ is preferred to $\beta \in y$ for all correlated alternative pairs (α, β). This implication is equivalent to that of the dominance axiom in the Anscombe–Aumann framework, except that the correlation between alternatives in menus x and y is determined by an endogenously derived set C, instead of exogenously given states of the world. Accordingly, Axiom 6 derives a state-*in*dependent utility function, that is, the utility level attained by each alternative in the menu depends only on its commitment ranking, and is independent of the realization of mental states, which sets our approach apart from existing studies. Note that, by the definition of random menus, this axiom also carries the inverse implication; that is, if $\beta \succsim \alpha$ for all $(\alpha, \beta) \in C$, then $y \succsim x$.

The next axiom renders the domain of choice sufficiently rich.

Axiom 7 (Richness of domain)
(a) There exist x, $y \in \mathcal{A}$ such that $x \succ y$.
(b) For all $n \in \mathbb{N}$, there exist $x_n^* = \{\beta_1^n, \cdots, \beta_n^n\}$ such that $\lambda x_n^* \oplus (1-\lambda)x_{n+1}^* \sim \cup_{i=1}^{n+1}\{\lambda \beta_{\min\{i,n\}}^n + (1-\lambda)\beta_i^{n+1}\}$ for all $n \in \mathbb{N}$ and $\lambda \in [0,1]$.
(c) For all $x = \{\beta_1, \cdots, \beta_n\} \in \mathcal{A}$, $1 \leq k \leq n$, and $\bar{\beta} \in \Delta(Z)$, there exists $x' = \{\beta_1', \cdots, \beta_n'\} \in \mathcal{A}$ such that $\beta_i' \sim \beta_i$ for all $i \neq k$, $\beta_k' \sim \bar{\beta}$, and $\lambda x \oplus (1-\lambda)x' \sim \cup_{i=1}^{n}\{\lambda \beta_i + (1-\lambda)\beta_i'\}$ for all $\lambda \in [0,1]$.

Condition (a) is the standard nondegeneracy axiom, whereas conditions (b) and (c) characterize the richness of the domain with respect to the position-dependent correlated choice. Condition (b) ensures the existence of menus x_1^*, x_2^*, $\cdots$, such that the alternative in each position of menu x_n^* will be correlated with its counterpart in x_{n+1}^* (i.e., β_i^n is mixed with β_i^{n+1} for $i = 1, \cdots, n$). This condition derives mental states that are consistently indexed for menus with different cardinalities. To interpret condition (c), suppose that $\beta_i' = \beta_i$ for all $i \neq k$, which gives $\lambda x \oplus (1-\lambda)x' \sim x \setminus \{\beta_k\} \cup \{\lambda \beta_k + (1-\lambda)\beta_k'\}$ for all $\lambda \in [0,1]$. This implies that alternative β_k in any (or the k-th) position of x will be correlated with alternative β_k' having an arbitrary commitment ranking, whereas alternatives β_i will be correlated with themselves for all $i \neq k$. This condition also allows for indifference between β_i' and β_i for all $i \neq k$, instead of equivalence between them.

Note that condition (c) derives a bijection from the set $\mathcal{A}_n$ of menus to that of Anscombe–Aumann acts, because correlation between alternatives is independent of the commitment ranking. Together with the obtained von Neumann–Morgenstern (vN-M) utility function, this generates a linear space that is crucial to our main theorem. However, if correlated choice is related to the commitment ranking, such a bijection cannot be obtained. For example, if alternatives with identical commitment rankings in each menu are correlated, that is, $\lambda x \oplus (1-\lambda) y \sim \{\lambda\alpha_i + (1-\lambda)\beta_i : i = 1, \cdots, n\}$ for all menus $x = \{\alpha_1, \cdots, \alpha_n\}$ and $y = \{\beta_1, \cdots, \beta_n\}$ such that $\alpha_1 \succsim \cdots \succsim \alpha_n$ and $\beta_1 \succsim \cdots \succsim \beta_n$, condition (c) is clearly violated, and only a mapping from $\mathcal{A}_n$ to a *subset* of Anscombe–Aumann acts can be obtained.

4 Representation theorem

This section states our main representation theorem. First, we define an ASC representation, which forms the primary focus of this study. We say that function $W : \Delta(\mathcal{A}) \to \Re$ *represents* $\succsim$ if $P \succsim Q$ whenever $W(P) \geq W(Q)$.

Definition 2 We call $(u, \phi, S, \mathcal{M})$ an *ASC representation of preference* $\succsim$ if the following conditions hold:

(a) $S = \cup_{n=1}^{\infty} S_n$, where $S_n = \{s_1, \cdots, s_n\}$ for all $n \in \mathbb{N}$;
(b) $\mathcal{M} = \cup_{n=1}^{\infty} \mathcal{M}_n$, where $\mathcal{M}_n \subseteq \Delta(S_n)$ is convex and closed for all $n \in \mathbb{N}$;
(c) $\phi : \mathcal{A} \times S \to \Delta(Z)$ is such that $\phi(x, s) \in x$ for all $x \in \mathcal{A}$ and $s \in S$;
(d) $V : \Delta(\mathcal{A}) \to \Re$ represents $\succsim$, where

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

for all $P \in \Delta(\mathcal{A})$, with an affine function $u : \Delta(Z) \to \Re$.

This representation is interpreted as follows: condition (a) defines the set S of mental states, while (b) defines the set $\mathcal{M}$ of probability measures over mental states. Condition (c) denotes the stochastic choice function ϕ that yields an alternative from a given menu $x \in \mathcal{A}$, depending on mental state $s \in S$,[8] and condition (d) denotes a representation V of $\succsim$ wherein a probability measure μ over mental states is chosen to maximize the expected utility using the vN-M utility function u. Note that the dependence of (1) on $\bar{n}(P)$, that is, the maximum cardinality of the menus in the support of P, captures a possible change in the evaluation of a random menu by a cardinality change.

For a nonrandom menu $x \in \mathcal{A}$, (1) can be simplified to:

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

[8] Peter Hammond suggests the alternative terminology of *selection function* for this type of choice function.

This representation implies that the DM chooses a probability measure μ over mental states to maximize her expected utility, considering the probability distributions over alternatives generated by μ and ϕ.

Next, the following regularity conditions are considered.

Definition 3 We refer to ϕ in an ASC representation $(u, \phi, S, \mathcal{M})$ as *regular* if the following conditions are satisfied:

(a) For all $x \in \mathcal{A}$, $\phi^x_{|x|} : S_{|x|} \to x$ defined by $\phi^x_{|x|}(\cdot) \equiv \phi(x, \cdot)$ is a bijection.
(b) For all $x \in \mathcal{A}$ and $n' \in \mathbb{N}$ such that $n' \geq |x|$, $\phi(x, s_{n'}) = \phi(x, s_{|x|})$.
(c) For all $n \in \mathbb{N}$ and $(\bar{\beta}_1, \cdots, \bar{\beta}_n) \in (\Delta(Z))^n$, there exists some $x \in \mathcal{A}$ such that $|x| = n$ and $\phi(x, s_i) \sim \bar{\beta}_i$ for $i = 1, \cdots, n$.

Further, we refer to an ASC representation $(u, \phi, S, \mathcal{M})$ as *regular* if ϕ is regular.

The first two conditions imply that mental states are position-relevant, whereas the final one renders mental states independent of the commitment ranking: condition (a) indicates that for all menus x with cardinality n, the stochastic choice function derives a bijection between the set S_n of mental states and menu x. That is, mental states correspond to each of n positions of alternatives in the menu, rather than the alternatives themselves; condition (b) implies that, for all natural numbers n' greater than the cardinality of menu x, mental state $s_{n'}$ triggers the choice of the "last" alternative in x, that is, $\phi(x, s_{|x|})$; condition (c) guarantees the existence of a menu such that the alternative triggered by each position-relevant mental state has an arbitrary commitment ranking. Note that condition (a) does not necessarily imply that all alternatives in any given menu are chosen with a positive probability, because there may exist $s \in S$ such that $\mu(s) = 0$ for all $\mu \in \mathcal{M}$.

Next, we need the following uniqueness concept to state our representation theorem. Given an ASC representation $(u, \phi, S, \mathcal{M})$ and $n \in \mathbb{N}$, we refer to a mental state $s \in S_n$ such that $\mu(s) > 0$ for some $\mu \in \mathcal{M}_n$ as being *relevant to* S_n.

Definition 4 We consider an ASC representation $(u, \phi, S, \mathcal{M})$ of $\succsim$ to be *essentially unique* if the following conditions hold for all ASC representations $(u', \phi', S', \mathcal{M}')$ of $\succsim$:

(a) u' is a positive affine transformation of u; that is, there exist $a > 0$ and $b \in \Re$ such that $u'(\cdot) = au(\cdot) + b$.
(b) For $n \in \mathbb{N}$ and the sets $\tilde{S}_n$ and $\tilde{S}'_n$ of mental states relevant to S_n and S'_n, respectively, there exists a bijection $\eta_n : \tilde{S}_n \to \tilde{S}'_n$ such that $\phi'(x, \eta_n(s))$ $= \phi(x, s)$ for all $x \in \mathcal{A}_n$ and $s \in \tilde{S}_n$.
(c) $\mathcal{M} = \cup_{n=1}^{\infty} \{\mu' \circ \eta_n \in \Delta(\tilde{S}_n) : \mu' \in \mathcal{M}'_n\}$ for η_n defined above.

Condition (a) is straightforward. Condition (b) implies that relevant mental states in S and S' trigger the choice of identical alternatives under some renumbering, whereas condition (c) renders the sets of probability measures over mental states equivalent after such a renumbering.

We now state the main theorem of this study.

Theorem 1 *Preference $\succsim$ satisfies Axioms 1–7 if and only if $\succsim$ admits a regular ASC representation $(u, \phi, S, \mathcal{M})$. Furthermore, the ASC representation is essentially unique.*

The proof of this theorem is provided in the appendix, while the necessity part is outlined in the following steps. First, for all menus x and y such that $|x| = |y|$, the perfectly correlated mixtures of menus axiom obtains a bijection τ from x to y that characterizes the DM's perception of random menus (Lemma 1). Then, together with the richness of domain axiom, we derive the set of mental states $S = \cup_{n=1}^{\infty}\{S_n\}$ and regular stochastic choice function ϕ (Lemma 2). By construction, all random menus P can be identified with menus of cardinality $\bar{n}(P)$.

Second, weak order, Archimedean continuity for menus with constant cardinalities, and S-independence derive a vN-M utility function u representing the commitment ranking $\succsim$ (Lemma 3). We also demonstrate the existence of a preference representation over finite menus (Lemma 4).

Third, we derive a cognitive control structure from S-independence and ARI, together with the set of mental states and the regular stochastic choice function obtained above. For a fixed $n \in \mathbb{N}$, the regularity of ϕ ensures that $\mathcal{A}_n$ is rich enough to mimic all the Anscombe–Aumann acts with state space S_n using functions $\phi_n^x(\cdot) \equiv \phi(x, \cdot)$.[9] By defining $K \equiv u(\Delta(Z))$, the set of composite functions $u \circ \phi_n^x : S_n \to K$ of utility function u and functions $\phi_n^x(\cdot)$ for all $x \in \mathcal{A}_n$ will generate a linear space B, endowed with mental-state-wise scalar multiplication and addition. Therefore, we derive the desired preference representation in a manner similar to Gilboa and Schmeidler's (1989) maxmin expected utility theory (MMEU), except that ARI has the inverse implication of uncertainty aversion, resulting in a representation that maximizes, rather than minimizes, the expected utility (Lemmas 5 and 6).[10]

Finally, we apply the same arguments to menus with all n cardinalities. The proof for the uniqueness result is similar to that given by Gilboa and Schmeidler.

A few remarks are in order. First, given a menu $x \in \mathcal{A}$ and a probability measure $\mu_x \in \mathcal{M}_{|x|}$ over mental states that attains the optimality in (2), we can define a *random choice rule (RCR)*, that is, a probability distribution over alternatives, as

$$\rho_x(\beta) = \mu_x(\{s : \phi(x, s) = \beta\})$$

[9] Chandrasekher (2015) independently developed a technique to apply the arguments in the Anscombe–Aumann framework to menu preferences by providing additional structures to the menus. A major difference in our approach is the imposition of perfectly correlated mixtures of menus, whereas Chandrasekher employed uncorrelated mixtures. This difference eventually results in representations with a vN-M utility function (combined with the stochastic choice function) in ASC, and a Strotzian value function in Chandrasekher's model.

[10] We discuss the implication of the preference representation obtained by reversing ARI's implication in Section 6.3.

for all $\beta \in x$. This RCR generally violates regularity (Luce and Suppes 1965) and WARSP (Bandyopadhyay et al. 1999), because μ_x depends on menu x.[11] The menu-dependence stems from the interplay between two factors—the correspondences between position-relevant mental states and alternatives in the menu, and the optimal choice of probability measure over mental states—and accommodates choice anomalies, such as those discussed in Section 6.1.

Second, ASC regards nonextreme points of a menu's convex hull differently from the random utility approach (e.g., Dekel et al. 2001). Consider degenerated probability distributions over prizes b_1, $b_2 \in Z$, and assume that $\beta \in \Delta(Z)$ generates each of b_1 and b_2 with a probability of 1/2. Then, β is a nonextreme point of the convex hull of $x = \{b_1, b_2, \beta\}$. As highlighted by Gul and Pesendorfer (2006), random utility models generally satisfy *extremeness*, that is, they allow only the extreme points of x's convex hull, namely, b_1 and b_2, to be chosen with a positive probability. Moreover, these models satisfy *indifference to randomization (IR)* (Dekel, et al. 2001), that is, menu x being indifferent to $y = \{b_1, b_2\}$, because they have identical convex hulls. In contrast, ASC generally violates both extremeness and IR: for menu x and the RCR ρ_x defined above, ASC allows for $\rho_x(\beta) > 0$ if, for example, all $\mu \in \mathcal{M}_3$ have full support. That is, the nonextreme alternative β can be chosen from x with a positive probability, and thus x will be distinguished from menu y, even if they have identical convex hulls. This property of ASC not only enables us to characterize stochastic choice with full support, such as trembling hands, but also produces sensitivity to adding, removing, or replacing alternatives in the menu, which is another driving force for explaining the choice anomalies in Section 6.1.

Third, in addition to characterizing stochastic choice for menus with a given cardinality, ASC can naturally relate stochastic choices over menus with different cardinalities using Bayes' rule. That is, under an additional axiom, the set $\mathcal{M}_m$ of probability measures over mental states can be obtained by conditioning each probability measure in $\mathcal{M}_n$ for all natural numbers $m < n$ (see Online Resource 1 for details). This result exploits the parallelism of ASC to MMEU, and is relevant to existing stochastic choice models. In particular, in the case of uncontrolled stochastic choice, that is, the sets $\mathcal{M}_n$ of probability measures over mental states consisting of singletons, the relative probability of one mental state over another will be independent of the menu, which carries an implication similar to that of Luce (1959).

Finally, although we have mainly discussed our model in terms of mistakes, it can also be interpreted as a form of temptation: suppose that the DM is a random Strotzian, that is, she cannot resist the temptation of choosing the triggered alternative once a mental state has been realized. However, she can still exercise self-control by regulating a probability distribution over mental states through some cognitive strategies, such as making a resolution or as-

[11] Our approach is in line with that of Noor and Takeoka (2010, 2015) in that both consider a menu preference that violates the Independence of Irrelevant Alternatives axiom and the Weak Axiom of Revealed Preference (WARP). However, Noor and Takeoka focus on the deterministic choice of alternatives from a menu, whereas we allow for stochastic choice.

signing herself a penalty/reward. In this case, $\mathcal{M}$ in ASC can be interpreted as the set of such cognitive strategies available to the DM. This interpretation is also relevant to two-stage models that have recently been considered in the context of menu preference (Ahn and Sarver 2013; Ergin and Sarver 2010, 2015; Nehring 2006), which we discuss in Section 6.4.

The essential uniqueness of an ASC representation renders the comparison of $\mathcal{M}$ meaningful. In the next section, we indicate that the size of $\mathcal{M}$ can be interpreted as an index of rationality. This section, however, concludes by discussing four special cases of ASC, each of which varies in $\mathcal{M}$.

First, consider $\mathcal{M}_n = \Delta(S_n)$ for all $n \in \mathbb{N}$, which corresponds to the exact maximization of state-independent utility, that is, $V(x) = \max_{\beta \in x} u(\beta)$. In this case, *all* probability measures over position-relevant mental states are available, and thus the DM can choose the alternative in any given position with a probability of one. Section 5.1 demonstrates that monotonicity both implies and is implied by this special case.

The second special case is $\mathcal{M}_n = \{\mu_n\}$ for all $n \in \mathbb{N}$ and some $\mu_n \in \Delta(S_n)$, which obtains $V(x) = \int_{S_{|x|}} u(\phi(x,s))d\mu_{|x|}(s)$. This corresponds to uncontrolled stochastic choice; that is, the DM has no control over mental states and the choice of alternative will be solely determined by a single probability measure μ_n (see Online Resource 1 for its axiomatization).

The third is $\mathcal{M}_n = \{(1-\epsilon)\mu + \epsilon\mu_n : \mu \in \Delta(S_n)\}$ for all $n \in \mathbb{N}$, and certain $\mu_n \in \Delta(S_n)$ and $\epsilon \in [0,1]$. In this case, all the probability measures over mental states are available with a probability of $1-\epsilon$, whereas stochastic choice is uncontrollable with a probability of ϵ. Eventually, we obtain the trembling-hand representation (Selten 1975) as follows:

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

which implies that the DM chooses the best alternative from the menu with a probability of $1-\epsilon$ and randomly selects alternatives with a probability of ϵ.[12][13]

Finally, a special case of ASC can be interpreted as a choice with limited attention (Manzini and Mariotti 2014, 2015; Masatlioglu et al. 2012; Wright and Barbour 1977). For all $n \in \mathbb{N}$, let $Q_n \subseteq S_n$ and $\mathcal{M}_n = \Delta(Q_n)$. Then, the resulting ASC representation is $V(x) = \max_{\beta \in \phi(x, Q_{|x|})} u(\beta)$. That is, Q_n corresponds to the set of attended positions of alternatives in the menu, from

[12] Although we do not provide a formal axiomatization of the trembling-hand ASC representation, for simplicity, it is derived by exploiting an axiomatization of ϵ-contamination (e.g., Nishimura and Ozaki 2006) and the similarity between ASC and MMEU.

[13] This representation is reminiscent of that of Chatterjee and Krishna (2009), who considered a DM that maximizes the normative utility with a probability of $1-\epsilon$ and the "alter-ego" utility with a probability of ϵ. Specifically, if we interpret the composite functions of u and ϕ as exactly maximized state-dependent utility functions, (3) implies that the DM maximizes the normative utility with a probability of $1-\epsilon$ and the *multiple* "alter-ego" utilities with a probability of ϵ. Unlike their study, however, our model permits nonextreme points of the menu's convex hull to be chosen with a positive probability in the "ϵ" event, which is crucial to interpreting it as the trembling-hand case.

which the DM chooses the best alternative; on the other hand, the alternatives associated with the mental states in $S_n \setminus Q_n$ will be disregarded, irrespective of their commitment rankings. In other words, the DM ex post fails to choose the alternatives associated with mental states in $S_n \setminus Q_n$ because of oversights or mistakes, although she is ex ante aware thereof.[14] Typically, $Q_n = \{s_1, \cdots, s_m\}$ for some $m \leq n$ implies that the DM's attention is confined to the first m alternatives of the n-cardinality menu, based on some criteria. Note that an ASC representation with limited attention permits stochastic choice as well as deterministic choice by allowing $\mathcal{M}_n$ to be a proper subset of $\Delta(Q_n)$. Moreover, in contrast to Manzini and Mariotti (2014, 2015), whose primary focus is on stochastic choice with limited attention, this special case can be derived from a general ASC representation axiomatized by Theorem 1.

5 Preferences for flexibility and commitment

This section explores the implications of ASC with regard to the sizes of menus, a topic frequently examined in the literature.

5.1 Preferences for flexibility

First, we discuss monotonicity, which is interpreted as a preference for flexibility. This axiom has been considered in many studies on menu preferences (e.g., Dekel et al. 2001; Ergin and Sarver 2010, 2015; Kreps 1979; Nehring 1999).

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 preference does not satisfy monotonicity unless the sets $\mathcal{M}_n$ in the representation comprise all probability measures over mental states for all natural numbers n.

Theorem 2 *Assume that* $\succsim$ *admits a regular ASC representation* $(u, \phi, S, \mathcal{M})$. *Then, the following statements are equivalent:*

(a) $\succsim$ *satisfies Axiom 8.*
(b) $\mathcal{M}_n = \Delta(S_n)$ *for all* $n \in \mathbb{N}$.

The proof is presented in the appendix. This theorem indicates that, despite their apparent similarity, no nontrivial intersection occurs between ASC and the models of a preference for flexibility, such as that of Dekel et al. (2001): if monotonicity is required, a regular ASC is reduced to the exact maximization of a state-*in*dependent utility function because, otherwise, adding undesirable alternatives may deteriorate the evaluation of a menu by increasing the chance of an inadvertent choice being made.[15]

[14] A similar "implementation error" interpretation of limited attention in the context of menu preference is discussed by Manzini and Mariotti (2015).

[15] A similar argument indicates that the set betweenness axiom introduced by Gul and Pesendorfer (2001) (i.e., $x \succsim y$ implying $x \succsim x \cup y \succsim y$) is equivalent to exact utility max-

5.2 Preferences for commitment

Next, we investigate the other extreme, that is, a preference for commitment.

Definition 5
(a) For all $x \in \mathcal{A}$, preference $\succsim$ exhibits *a preference for commitment to a singleton menu at* x if there exists $\beta \in x$ such that $\{\beta\} \succ x$.
(b) *Preference* $\succsim_2$ *exhibits a preference for commitment to a singleton menu greater than preference* $\succsim_1$ if, for all $x \in \mathcal{A}$ and $\beta \in x$,

$$\{\beta\} \succ_1 x \Rightarrow \{\beta\} \succ_2 x.$$

Preferences for commitment to a singleton menu are stronger than the preferences for commitment defined by Gul and Pesendorfer (2001). Dekel and Lipman (2012) and Ergin and Sarver (2010) conducted similar interpersonal comparisons.[16] The following observation indicates that ASC generally exhibits a preference for commitment to a singleton menu unless an exact utility maximization is possible.

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

To understand the rationale behind this implication, let $\hat{\beta}$ be the alternative representing the highest commitment ranking in menu x such that no probability measure in $\mathcal{M}_{|x|}$ assigns probability one to the mental state that triggers the choice of $\hat{\beta}$ (the existence of such x and $\hat{\beta}$ follows from regularity condition (c) and $\mathcal{M} \subsetneq \cup_{n=1}^{\infty} \Delta(S_n)$). The resulting ASC preference clearly implies that $\{\hat{\beta}\} \succ x$.

Next, the following theorem relates the interpersonal comparison of a preference for commitment to the size of the set $\mathcal{M}$ of probability measures over mental states.

Theorem 3 *Assume that DM$_i$'s preference* $\succsim_i$ *admits a regular ASC representation* $(u, \phi, S, \mathcal{M}^i)$ *for* $i = 1$ *and* 2. *As a result,* $\succsim_2$ *exhibits a preference for commitment to a singleton menu greater than that of* $\succsim_1$ *if and only if* $\mathcal{M}^1 \supseteq \mathcal{M}^2$.

imization for a regular ASC: the latter trivially implies the former. Conversely, under the lemma proved by Gul and Pesendorfer, set betweenness implies in ASC that only the mental states associated with the normatively best and worst alternatives in menu x are assigned a positive probability by the optimal probability measure over mental states chosen in (2). Moreover, the normatively worst alternative is assigned a zero probability; otherwise, regularity condition (c) can replace the alternative, so that the optimal probability measure over mental states assigns a positive probability to an alternative that is neither normatively best nor normatively worst in the menu, which contradicts the implication of Gul and Pesendorfer's lemma. Thus, $\mathcal{M}_{|x|}$ includes a probability measure that assigns a probability of one to an arbitrary mental state. It follows from the closedness and convexity that $\mathcal{M}_{|x|}$ comprises all the probability measures over mental states.

[16] A condition similar to Definition 5a can also be found in the desire for commitment axiom proposed by Dekel et al. (2009).

The proof is given in the appendix. Intuitively, the size of $\mathcal{M}$ can be interpreted as an index of *rationality*: for DMs possessing identical commitment rankings, the smaller $\mathcal{M}$ becomes, the more restrictive is the DM's choice of probability measure over mental states. Thus, $\mathcal{M}^2$ being smaller than $\mathcal{M}^1$ implies that DM_2 is more likely to make an erroneous choice from the menu than DM_1, which means that DM_2 has a higher preference for commitment to a singleton menu than DM_1. Moreover, if we accept the temptation interpretation of ASC discussed in Section 4, Theorem 3 implies that DM_2 is more prone to temptation generated by ex ante suboptimal alternatives in the menu than DM_1, because a smaller number of cognitive strategies for suppressing temptation are available to the former than to the latter. Finally, this result is also compatible with the psychology literature that associates higher cognitive abilities with lower vulnerability to psychological effects such as the attraction and position effects (Krueger and Salthouse 2011; Sherman et al. 2008; Tentori et al. 2001) because, as indicated in Section 6.1, ASC explains such choice anomalies by restricting $\mathcal{M}$. Note that comparing the size of $\mathcal{M}$ is meaningful in our model, because a regular ASC representation is essentially unique.

In the trembling-hand case (3), the above set inclusion can be described by the single parameter ϵ. Let $\mathcal{M}_n = \{(1-\epsilon)\mu + \epsilon\mu_n : \mu \in \Delta(S_n)\}$ and $\mathcal{M}'_n = \{(1-\epsilon')\mu + \epsilon'\mu_n : \mu \in \Delta(S_n)\}$ for some $\mu_n \in \Delta(S_n)$ and ϵ, $\epsilon' \in [0,1]$. It follows that we have $\epsilon \leq \epsilon'$ if and only if $\mathcal{M}_n \supseteq \mathcal{M}'_n$. This provides a behavioral foundation for the classic argument proposed by Selten (1975), who interpreted ϵ as an index of the DM's rationality.

This theorem parallels that proposed by Ghirardato and Marinacci (2002, Theorem 17), who defined an index of comparative ambiguity aversion by the set of priors in MMEU, except that the contraction of the set of priors implies a *lower* preference for certainty in MMEU, whereas the contraction of the set of probability measures over mental states implies a *greater* preference for commitment to a singleton menu in ASC.

6 Discussion

6.1 Choice anomalies

This subsection demonstrates that ASC accommodates certain choice anomalies reported in the literature, which reflects the following two key factors. First, these choice anomalies are generally caused by a position change among alternatives in the menu as a result of adding, removing, or replacing an alternative. ASC captures this effect by varying the connections between position-relevant mental states and alternatives, while holding the alternatives' commitment rankings unchanged.

Second, because the set $\mathcal{M}$ of probability measures over mental states can be interpreted as an index of rationality, as discussed in the previous section, ASC attributes nonstandard choice patterns to a restriction on $\mathcal{M}$, rather than the violation of axioms such as WARP, WARSP, and transitivity. This provides a unified framework that regards both exact utility maximization and choice anomalies as special cases of ASC, and consistently relates stochastic and deterministic choices. To highlight this property, the following discussion focuses on stochastic, rather than deterministic, choice, but deterministic choice can also be obtained by allowing ϵ and ϵ' below to be zero.

In the following, for simplicity, we write $c(x) = (\beta_1, \cdots, \beta_{|x|})$ for all $x \in \mathcal{A}$ to imply $\phi(x, s_i) = \beta_i$ for $i = 1, \cdots, |x|$.

6.1.1 Attraction effect

The attraction effect occurs when adding an unchosen alternative to an existing menu changes the choice of alternative in that menu (e.g., Huber et al. 1982). Let α, β, and $\delta_\beta \in \Delta(Z)$. Assume that α 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. Eventually, the DM chooses α from $x = \{\alpha, \beta\}$ and β from $y = \{\alpha, \beta, \delta_\beta\}$ with the highest probability (often a probability of one). This example clearly violates WAR(S)P.

We can easily explain this behavioral pattern in ASC by assuming that $\alpha \succ \beta \succ \delta_\beta$, $c(y) = (\beta, \delta_\beta, \alpha)$, $\mathcal{M}_2 = \{\mu \in \Delta(S_2) : \mu(s_i) \leq 1-\epsilon \text{ for } i = 1, 2\}$, and $\mathcal{M}_3 = \{\mu \in \Delta(S_3) : \mu(s_i) \leq 1 - \epsilon \text{ for } i = 1, 2 \text{ and } \mu(s_3) \leq \epsilon'\}$ for sufficiently small ϵ and ϵ'.[17] That is, despite α having the highest commitment ranking of the three alternatives (and being chosen from x with a probability of $1 - \epsilon$), it is chosen from y with a very low probability (equal to ϵ'), because the addition of δ_β drives α into a "blind spot," that is, the least attended position, which is associated with mental state s_3.[18]

6.1.2 Cyclical choice

ASC can also explain cyclical choice. For some α, β, $\gamma \in \Delta(Z)$, assume that $c(\{\alpha, \beta\}) = (\alpha, \beta)$, $c(\{\beta, \gamma\}) = (\beta, \gamma)$, $c(\{\gamma, \alpha\}) = (\gamma, \alpha)$, and $\mathcal{M}_2 = \{\mu \in \Delta(S_2) : \mu(s_1) = 1-\epsilon\}$ for a sufficiently small ϵ. Then, α is chosen from $\{\alpha, \beta\}$,

[17] The reader may suspect that the above argument relies on the fact that x and y have different cardinalities, and are thus evaluated by different sets of probability measures over mental states. However, ASC also accommodates similar choice anomalies for menus of identical cardinality. For example, Tyszka (1983) reported that, for α, β, γ, and $\delta \in \Delta(Z)$, the majority of the participants choose α from $x = \{\alpha, \beta, \gamma\}$, despite choosing β from $y = \{\alpha, \beta, \delta\}$. Assuming that $\alpha \succ \beta \succ \gamma \sim \delta$, $c(x) = (\alpha, \gamma, \beta)$, $c(y) = (\beta, \delta, \alpha)$, and $\mathcal{M}_3 = \{\mu \in \Delta(S_3) : \mu(s_i) \leq 1 - \epsilon \text{ for } i = 1, 2 \text{ and } \mu(s_3) \leq \epsilon'\}$ for sufficiently small ϵ and ϵ', the above choice pattern is consistent with the ASC preference.

[18] This result does *not* imply that the attraction effect *generally* accompanies a preference for commitment. However, the attraction effect (and other choice anomalies) can be associated with a *specific* type of preference for commitment: in this example, the DM would naturally prefer menu x to y, because she is ex ante aware that decoy δ_β attracts her attention to the suboptimal alternative β, and is thus willing to exclude δ_β from the menu. Another interpretation is that the menu choice is made by the social planner, who prefers to exclude suboptimal alternatives from the menu to prevent the DM from making an erroneous alternative choice. A discussion relevant to this interpretation can be found in Manzini and Mariotti (2015).

β is chosen from $\{\beta, \gamma\}$, and γ is chosen from $\{\gamma, \alpha\}$ with probability $1-\epsilon$. That is, cyclical choice occurs because the alternative that is (objectively or subjectively) positioned first in each pair is chosen with a higher probability. This indicates that the stochastic choice function ϕ (or equivalently, function c) and the sets $\mathcal{M}_n$ of probability measures over mental states may impact the choice of alternatives more significantly than the alternatives' commitment rankings would.

Specifically, consider the case $\epsilon = 0$. This implies that only the alternative associated with a specific position (or mental state) is chosen with a probability of one, regardless of its commitment ranking, and carries an implication similar to that of Strotz (1955) and Gul and Pesendorfer's (2001) overwhelming temptation.

6.1.3 Position effects

Various position effects suggest that alternatives located at the beginning, end, or in other specific positions (typically the middle) of the menu have a higher probability of being chosen than others (e.g., Bruine de Bruin 2005; Christenfeld 1995; Murdock 1962). To explain these effects in ASC, we assume that $c(x) = (\alpha, \beta, \gamma)$ for a menu $x = \{\alpha, \beta, \gamma\}$. That is, alternatives α, β, and γ are physically ordered first, second, and third in the menu, and mental state s_i triggers the choice of the alternative in the i-th position for $i = 1$, 2, and 3. We also assume uncontrolled stochastic choice, i.e., $\mathcal{M}_3 = \{\mu\}$, for simplicity. Then, the probability of each alternative being chosen is solely determined by the probability measure μ over mental states: $\mu(s_1) > \mu(s_2) \geq \mu(s_3)$ implies the primacy effect, that is, the first item α is chosen with the highest probability; $\mu(s_1) \leq \mu(s_2) < \mu(s_3)$ implies the recency effect, that is, the last item γ is chosen with the highest probability; $\mu(s_1) < \mu(s_2)$ and $\mu(s_2) > \mu(s_3)$ imply that the alternative β in the middle is chosen with the highest probability.

6.2 Mixtures of menus

As we have noted, ASC is in stark contrast to the random utility approach in the menu preference framework (Ahn and Sarver 2013; Chatterjee and Krishna 2009, 2012; Dekel and Lipman 2012; Dekel et al. 2001, 2009; Ergin and Sarver 2010, 2015; Gul and Pesendorfer 2001; Kreps, 1979; Nehring 1999; Noor and Takeoka 2010, 2015; Sarver 2008). Accordingly, this subsection discusses the relationship between the axioms that are crucial to the difference, that is, Axioms 2 and 2' proposed in Section 3.

First, the following observation highlights the difference between the two axioms.

Observation 2 *Axiom 2' is equivalent to the following axiom:*

Axiom 2" (Uncorrelated mixtures of menus, another representation)

(a) For all $x \in \mathcal{A}$ and $\lambda \in [0,1]$, $\lambda x \oplus (1-\lambda)x \sim \lambda x + (1-\lambda)x$.
(b) $\succsim$ has consistent perceptions of random menus.

The proof is given in the appendix. Because Axioms 2b and 2"b are equivalent, this observation indicates that the apparent difference between Axioms 2 and 2' reduces to the difference between Axioms 2a and 2"a, that is, whether a randomization between *identical* menus is indifferent to the original menu or the Minkowski sum, which can readily be tested. On the other hand, Axioms 2b and 2"b, or consistent perceptions of random menus, create a common ground to characterize preferences that evaluate random menus differently.

Next, we discuss the difference between the two axioms with respect to the sensitivity to a change of membership, such as adding, deleting, or replacing an alternative in the menu. As argued by Dekel et al. (2001), iteratively applying Axiom 2' to the identical menu x, along with independence (or relevant axioms) and continuity, implies the IR property, that is, any menu x being indifferent to its convex hull. This property derives random utility (i.e., the exact maximization of the state-dependent utility), and a change of membership does not affect the evaluation of the menu unless the menu's convex hull is altered. In contrast, Axiom 2, which is imposed in ASC, identifies a randomization between identical menus x with the menu x itself. Thus, even the iterative application of the axiom does not render menu x indifferent to its convex hull, which invalidates IR. This also derives a preference representation with a stochastic choice function (i.e., a probability measure over mental states or alternatives) that is possibly full support as in trembling hands and sensitive to a membership change of the menu, the latter of which is crucial to accommodate the choice anomalies in Section 6.1.

6.3 Self vs. nature

As we have demonstrated, the DM chooses a probability measure over mental states to *maximize* her ex ante expected utility in ASC. Ergin and Sarver (2010, 2015) considered a model similar to ours, wherein a probability measure over subjective states was chosen to maximize the expected utility. In contrast, Gilboa and Schmeidler (1989) modeled ambiguity aversion using a probability measure over the set of exogenous states that *minimizes* the expected utility, while Epstein et al. (2007) extended this approach to endogenous states. Ahn (2007), Olszewski (2007), and Chatterjee and Krishna (2012) endorsed the latter view by adopting another approach.

The difference between the maximization and minimization approaches can be summarized as whether *self* or *nature* chooses a probability measure over states: if the DM has control over her subjective or mental states, she can presumably maximize the expected utility by choosing a probability measure over states. However, if the ex post choice of alternative depends on the states chosen by nature, such as weather, natural disasters, and economic booms and recessions, the DM may only be concerned with the minimum possible expected utility, because the realization of these states is beyond her control.

If we reverse the implication of ARI, while retaining the other axioms, the resulting set of axioms yields the counterpart of an ASC representation; in this counterpart, a probability measure on the set of states is chosen to *minimize*, rather than maximize, the expected utility. However, the above discussion indicates that the implication of this model differs significantly from that of ours, because nature or other conditions beyond the DM's control, and not the DM herself, are assumed to set the probability measure.[19]

6.4 Related literature

This subsection reviews the relevant literature. First, several two-stage models relevant to ASC have recently been considered. Ahn and Sarver (2013) characterized the first-stage menu preference to uniquely identify the second-stage stochastic choice generated by random utility. Unlike ASC, however, they did not derive a cognitive-control structure. On the other hand, Ergin and Sarver (2010, 2015) considered a model similar to cognitive control, wherein the DM maximizes the expected utility by choosing some interim action that affects the second-stage preference. In particular, Ergin and Sarver (2015) assumed the domain of random menus, as we have, to derive a uniqueness result. As noted in Section 6.2, however, this line of study implicitly or explicitly imposes Axiom 2', whereas ASC imposes Axiom 2. Thus, these two approaches are different with respect to their interpretations (i.e., random utility vs. stochastic choice) as well as perceptions of random menus. Nehring (2006) considered a model wherein the DM chooses the optimal preference over the alternatives before an alternative is chosen from a menu. Unlike this model, however, ASC allows for stochastic choice and is essentially unique.

Second, the interplay between the ex ante optimality and ex post choice considered in this study resembles the dichotomy of *deliberative* and *affective* systems (Loewenstein and O'Donoghue 2004). Similar dichotomies prevail in economics (Chatterjee and Krishna 2009; Fudenberg and Levine 2006; Gul and Pesendorfer 2001; Kahneman et al. 1997; Thaler and Shefrin 1981), as well as in psychology and neuroscience. Particularly, the literature relevant to cognitive control and higher cognitive abilities reducing vulnerability to various psychological effects (Gross 1998; Krueger and Salthouse 2011; Larsen 2000; Sherman et al. 2008; Tentori et al. 2001) generally assumes such a model. From this perspective, ASC can be interpreted as an affective system triggering ex post stochastic choice and a deliberative system maximizing the expected utility by cognitive control.

[19] Nehring's (1999) argument parallels the relationship between the maximization and minimization approaches, as he used concave capacities to denote utility maximization. In contrast, Schmeidler (1989) used convex capacities to model ambiguity aversion.

7 Concluding remarks

This study has mainly characterized mistakes driven by position-dependent choice in the context of menu preference. Our ASC model also includes a range of subclasses relevant to the existing literature and choice anomalies.

Selten (1975), who first identified stochastic choice with bounded rationality, 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 argued that the probability distribution over alternatives generated by the DM's possible mistakes is determined by "some unspecified psychological mechanism."

This study contributes to the literature in terms of a decision theoretic foundation that unravels this "unspecified psychological mechanism": focusing on correlated choice triggered by position-relevant mental states, we have derived subjective stochastic choice from a preference over menus, rather than assuming an exogenous randomizing device over alternatives. The interpersonal comparison with respect to the sets of probability measures over mental states also provides a foundation for Selten's view that the error rate ϵ in the trembling-hand model can be considered as an index of rationality. Thus, we believe that our approach offers new insight into stochastic choice behavior.

Appendix

Proof of Theorem 1

For the sufficiency part of Theorem 1 (a regular ASC representation implies the axioms), we only present proofs for Axioms 2, 7b, and 7c, because the sufficiency for the other axioms is straightforward.

To confirm that Axiom 2 holds, we note that the affineness of u implies that $\int_{\mathcal{A}} u(\phi(x,s))\,dP(x) = \lambda_1 u(\phi(x_1,s)) + \cdots + \lambda_m u(\phi(x_m,s))$ for all finite random menus $P = (\lambda_1, x_1; \cdots; \lambda_m, x_m)$ that generate menu x_i with probability λ_i for $i = 1, \cdots, m$, and all $s \in S$. Then, Axiom 2b follows from regularity conditions (a) and (b), and considering $P = (\lambda, x; 1-\lambda, x)$ also implies Axiom 2a.

On the other hand, Axioms 7b and 7c are implied by the regularity conditions of ϕ. First, Axiom 7b is proved by induction. Let $x_1^* = \{\beta_1^1\}$ for an arbitrary $\beta_1^1 \in \Delta(Z)$. Next, fix $n \geq 1$ and assume that x_i^* satisfies the implication of Axiom 7b for $i = 1, \cdots, n$. The construction of ϕ implies, for any given $x_{n+1}^* \in \mathcal{A}$ such that $|x_{n+1}^*| = n+1$, that $\lambda x_n^* \oplus (1-\lambda) x_{n+1}^* \sim \cup_{i=1}^{n+1} \{\lambda \phi(x_n^*, s_i) + (1-\lambda)\phi(x_{n+1}^*, s_i)\}$ for all $\lambda \in [0,1]$. By defining $\beta_i^{n+1} \equiv \phi(x_{n+1}^*, s_i)$ for $i = 1, \cdots, n+1$, regularity conditions (a) and (b) imply that $x_{n+1}^* = \{\beta_1^{n+1}, \cdots, \beta_{n+1}^{n+1}\}$ and $\lambda x_n^* \oplus (1-\lambda) x_{n+1}^* \sim \cup_{i=1}^{n+1} \{\lambda \beta_{\min\{i,n\}}^n + (1-\lambda)\beta_i^{n+1}\}$ for all $\lambda \in [0,1]$. Because the same argument holds for all $n \geq 1$, we obtain Axiom 7b. Second, regularity condition (c) implies that, for all $x \in \mathcal{A}$,

$1 \le k \le |x|$, and $\bar{\beta} \in \Delta(Z)$, there exists $x' \in \mathcal{A}$ such that $|x'| = |x|$, $\phi(x', s_i) \sim \phi(x, s_i)$ for all $i \neq k$, and $\phi(x', s_k) \sim \bar{\beta}$, which implies Axiom 7c.

Below, we prove the necessity part of the theorem. The first lemma indicates that Axiom 2 derives a bijection between menus of identical cardinality that describes the DM's perception of random menus.

Lemma 1 *Axiom 2 implies the following axiom:*

Axiom 2* *For all x, $y \in \mathcal{A}$ such that $|x| = |y|$, there exists a bijection $\tau : x \to y$ such that $\lambda x \oplus (1-\lambda)y \sim \{\lambda\alpha + (1-\lambda)\tau(\alpha) : \alpha \in x\}$ for all $\lambda \in [0,1]$.*

Proof Axiom 2b (specifically, Definition 1a) implies that, for given x, $y \in \mathcal{A}$ such that $|x| = |y|$, there exists $C \subseteq x \times y$ such that $\lambda x \oplus (1-\lambda)y \sim \{\lambda\alpha + (1-\lambda)\beta : (\alpha, \beta) \in C\}$ for all $\lambda \in [0,1]$. By construction, we also have $\lambda' y \oplus (1-\lambda')x \sim \{\lambda'\beta + (1-\lambda')\alpha : (\alpha, \beta) \in C\}$ for all $\lambda' \in [0,1]$. From Axiom 2b (Definition 1b), it follows that

$$\lambda'' x \oplus (1-\lambda'')x \sim \{\lambda''\alpha + (1-\lambda'')\alpha' : (\alpha, \beta) \in C \text{ and } (\alpha', \beta) \in C\} \tag{4}$$

for all $\lambda'' \in [0,1]$. Axiom 2a implies that there exists some C such that the right-hand side of (4) is equivalent to x.

Now, assume that for all C satisfying (4), there is no bijection $\tau : x \to y$, such that $\beta = \tau(\alpha)$ if and only if $(\alpha, \beta) \in C$. This implies that (i) there exists $\tilde{\alpha} \in x$ such that $(\tilde{\alpha}, \beta) \notin C$ for all $\beta \in y$, (ii) there exist distinct $\tilde{\alpha}$, $\hat{\alpha} \in x$ such that $(\tilde{\alpha}, \beta)$, $(\hat{\alpha}, \beta) \in C$ for some $\beta \in y$, or (iii) there exist distinct $\tilde{\beta}$, $\hat{\beta} \in y$ such that $(\alpha, \tilde{\beta})$, $(\alpha, \hat{\beta}) \in C$ for some $\alpha \in x$. In case (i), we assume without loss of generality that there is no $C' \subseteq x \times y$ such that $C' \supsetneq C$ satisfies (4) in replacing C. Then, the right-hand side of (4) cannot be indifferent to x for all $\lambda'' \in [0,1]$ because it never includes $\tilde{\alpha}$. In case (ii), we assume without loss of generality that there is no $C' \subseteq x \times y$ such that $C' \subsetneq C$ satisfies (4) in replacing C. Then, the right-hand side of (4) cannot be indifferent to x for all $\lambda'' \in [0,1]$, because it includes $\lambda\tilde{\alpha} + (1-\lambda)\hat{\alpha}$. Finally, case (iii) is equivalent to case (ii) after exchanging the roles of x and y. Accordingly, all the cases contradict Axiom 2a.

The above discussion holds for all such menus x and y, and thus, we obtain the desired result. □

Next, we construct the sets of mental states and a regular choice function.

Lemma 2 *There exist $S_n = \{s_1, \cdots, s_n\}$, $S = \cup_{n=1}^{\infty}\{S_n\}$, and a regular ϕ such that*

$$\lambda x \oplus (1-\lambda)y \sim \cup_{s \in S_{\max\{|x|,|y|\}}}\{\lambda\phi(x, s) + (1-\lambda)\phi(y, s)\} \tag{5}$$

for all x, $y \in \mathcal{A}$ and $\lambda \in [0,1]$.

Proof We prove the lemma by induction. First, define $S_1 = \{s_1\}$ and $\phi(x, s_1) = \beta$ for all $x = \{\beta\} \in \mathcal{A}_1$. These definitions, combined with Lemma 1, imply regularity condition (a) and (5) for all singleton menus x and y.

Next, set $n \geq 1$ and assume that S_n and ϕ satisfy regularity condition (a) and (5) for all $x, y \in \mathcal{A}_n$. Axiom 7b implies that, for menus $x_n^* = \{\beta_1^n, \cdots, \beta_n^n\}$ and $x_{n+1}^* = \{\beta_1^{n+1}, \cdots, \beta_{n+1}^{n+1}\}$ defined in the axiom, $\lambda x_n^* \oplus (1-\lambda)x_{n+1}^* \sim \cup_{i=1}^{n+1}\{\lambda\beta_{\min\{i,n\}}^n + (1-\lambda)\beta_i^{n+1}\}$ for all $\lambda \in [0,1]$. Accordingly, by defining $S_{n+1} \equiv S_n \cup \{s_{n+1}\}$, $\phi(x_n^*, s_i) \equiv \beta_{\min\{i,n\}}^n$, and $\phi(x_{n+1}^*, s_i) \equiv \beta_i^{n+1}$ for $i = 1, \cdots, n+1$, we obtain $\lambda x_n^* \oplus (1-\lambda)x_{n+1}^* \sim \cup_{s\in S_{n+1}}\{\lambda\phi(x_n^*, s) + (1-\lambda)\phi(x_{n+1}^*, s)\}$ for all $\lambda \in [0,1]$ and $\phi(x_n^*, s_{n+1}) = \phi(x_n^*, s_n)$.

Now, iteratively applying Axiom 2b (i.e., $\succsim$ having consistent perceptions of random menus) to the above argument implies that, for all $m \leq n+1$, $\lambda x_m^* \oplus (1-\lambda)x_{n+1}^* \sim \cup_{s\in S_{n+1}}\{\lambda\phi(x_m^*, s) + (1-\lambda)\phi(x_{n+1}^*, s)\}$ for all $\lambda \in [0,1]$ and $\phi(x_m^*, s_i) = \phi(x_m^*, s_m)$ for $i = m+1, \cdots, n+1$. Moreover, Lemma 1 implies that, for all x, $y \in \mathcal{A}$ such that $|x| = m$ and $|y| = n+1$, there exist bijections $\hat{\tau} : x_m^* \to x$ and $\tilde{\tau} : x_{n+1}^* \to y$ such that $\lambda' x_m^* \oplus (1-\lambda')x \sim \{\lambda'\beta_i^m + (1-\lambda')\hat{\tau}(\beta_i^m) : i = 1, \cdots, m\}$ for all $\lambda' \in [0,1]$ and $\lambda'' x_{n+1}^* \oplus (1-\lambda'')y \sim \{\lambda''\beta_i^{n+1} + (1-\lambda'')\tilde{\tau}(\beta_i^{n+1}) : i = 1, \cdots, n+1\}$ for all $\lambda'' \in [0,1]$. Accordingly, by defining $\phi(x, s_i) = \hat{\tau}(\beta_{\min\{i,m\}}^m)$ and $\phi(y, s_i) = \tilde{\tau}(\beta_i^{n+1})$ for $i = 1, \cdots, n+1$, Axiom 2b implies that $\lambda x \oplus (1-\lambda)y \sim \cup_{s\in S_{n+1}}\{\lambda\phi(x, s) + (1-\lambda)\phi(y, s)\}$ for all $\lambda \in [0,1]$, and $\phi(x, s_i) = \phi(x, s_m)$ for $i = m+1, \cdots, n+1$. This implies (5) for all x, $y \in \mathcal{A}_{n+1}$. The construction of $\phi(y, \cdot)$ also gives regularity condition (a) for all $y \in \mathcal{A}_{n+1}$. Moreover, because the above argument implies that $\phi(x, s_{n'}) = \phi(x, s_{|x|})$ for all $x \in \mathcal{A}$ and $n' \in \mathbb{N}$ such that $n' \geq |x|$, regularity condition (b) is obtained.

Finally, by the construction of ϕ, Axiom 7c implies that, for all $1 \leq k \leq n$, $\bar{\beta} \in \Delta(Z)$, and $x = \{\beta_1, \cdots, \beta_n\}$ such that $\phi(x, s_i) = \beta_i$ for $i = 1, \cdots, n$, there exists $x' = \{\beta_1', \cdots, \beta_n'\}$ such that $\phi(x', s_i) = \beta_i'$ for $i = 1, \cdots, n$, $\beta_i' \sim \beta_i$ for all $i \neq k$, and $\beta_k' \sim \bar{\beta}$. Regularity condition (c) is obtained by iteratively applying the argument for $k = 1, \cdots, n$. □

The iterative application of (5) in Lemma 2 implies that a finite random menu $P = (\lambda_1, x_1; \cdots; \lambda_m, x_m)$ is indifferent to a finite nonrandom menu $\Psi(P) \equiv \cup_{s\in S_{\bar{n}(P)}}\{\lambda_1\phi(x_1, s) + \cdots + \lambda_m\phi(x_m, s)\}$. Thus, we focus on the preference over finite menus, rather than that over random menus.

Next, we show that there is an affine function u (i.e., $u(\lambda\beta + (1-\lambda)\gamma) = \lambda u(\beta) + (1-\lambda)u(\gamma)$ for all $\lambda \in [0,1]$ and β, $\gamma \in \Delta(Z)$) representing the commitment ranking.

Lemma 3 *There is an affine function u that represents $\succsim$ on $\Delta(Z)$; that is, for all β, $\beta' \in \Delta(Z)$, $u(\beta) \geq u(\beta')$ whenever $\{\beta\} \succsim \{\beta'\}$. Furthermore, u is unique up to a positive affine transformation.*

Proof The conclusion is straightforward from Axioms 1, 3, and 4. Note that the independence axiom for alternatives follows from singleton independence. □

The following lemma indicates that there is a functional J representing the restriction of $\succsim$ on $\mathcal{A}$.

Lemma 4 *Let u be an affine function derived from Lemma 3. Then, there is a functional $J : \mathcal{A} \to \Re$ such that*

(a) for all x, $y \in \mathcal{A}$, $x \succsim y$ if and only if $J(x) \geq J(y)$, and
(b) for all $\beta \in \Delta(Z)$ and $x = \{\beta\}$, $J(x) = u(\beta)$.

Proof For singleton menus, J is uniquely defined by (b). To define J for all menus, fix $\bar{\beta}$ and $\underline{\beta}$ such that $\{\bar{\beta}\} \succsim x \succsim \{\underline{\beta}\}$ for all $x \in \mathcal{A}$. It follows from Axiom 3 and Lemma 1 that there exists a unique $\lambda \in [0, 1]$ such that $x \sim \lambda\{\bar{\beta}\} \oplus (1-\lambda)\{\underline{\beta}\} \sim \{\lambda\bar{\beta} + (1-\lambda)\underline{\beta}\}$. Thus, by defining $J(x) = J(\{\lambda\bar{\beta} + (1-\lambda)\underline{\beta}\})$, J also satisfies (a). $\square$

Now, fix the cardinality n of the menus and the function u such that $u(\beta) > 1$ and $u(\beta') < -1$ for some β, $\beta' \in \Delta(Z)$ (which is allowed by Axiom 7a). We denote by B the linear space of all functions $a : S_n \to \Re$ endowed with state-wise scalar multiplication and addition; that is, for all a, $b \in B$ and $\lambda \in \Re$, λa and $a + b$ are defined by $(\lambda a)(s) = \lambda(a(s))$ and $(a + b)(s) = a(s) + b(s)$ for all $s \in S_n$, respectively. Because S_n is finite, B is equivalent to the linear subspace of simple functions, which we denote by B_0. We also define $K = u(\Delta(Z))$ and denote by $B_0(K)$ the set of simple functions that have range K. For all $\xi \in \Re$, we denote by $\xi^* \in B$ a constant function such that $\xi^*(s) = \xi$ for all $s \in S_n$. Now, define $\phi_n^x(\cdot) \equiv \phi(x, \cdot)$ on S_n. From regularity condition (c) of ϕ, which we proved in Lemma 2, it follows that the set of functions $u \circ \phi_n^x$ with x ranging over $\mathcal{A}_n$ is equivalent to $B_0(K)$, i.e., $\{u \circ \phi_n^x : x \in \mathcal{A}_n\} = B_0(K)$.

Let $\tilde{u} : \mathcal{A}_n \to K^{S_n}$ be such that $\tilde{u}(x) = u \circ \phi_n^x$ for all $x \in \mathcal{A}_n$. The following lemmas, which are counterparts of Gilboa and Schmeidler's (1989) lemmas, characterize the functional I on B derived from the axioms. Note that, unlike Gilboa and Schmeidler's argument, I is sublinear instead of superlinear and is denoted by the maximum of integrals rather than the minimum.

Lemma 5 *There is a functional $I : B \to \Re$ such that*

(a) for all $x \in \mathcal{A}$ such that $|x| = n$, $I(\tilde{u}(x)) = J(x)$,
(b) I is monotonic, that is, $a \geq b$ implies $I(a) \geq I(b)$ for all a, $b \in B$,
(c) I is sublinear (subadditive and homogeneous of degree one), and
(d) I is C-independent, that is, $I(a + \xi^*) = I(a) + I(\xi^*)$ for all $a \in B$ and $\xi \in \Re$.

Proof We define I on $B_0(K)$ by condition (a). This also implies that $I(\tilde{u}(\{\beta\})) = J(\{\beta\}) = u(\beta)$ for all $\beta \in \Delta(Z)$, and thus, $I(1^*) = 1$. The monotonicity (b) of I follows from Lemma 2 and Axiom 6. We indicate that I satisfies (c) and (d).

First, we show $I(\lambda a) = \lambda I(a)$ for all $\lambda \in (0, 1]$ and a, $\lambda a \in B_0(K)$. Let $y \in \mathcal{A}$ be such that $|y| = n$ and $a = \tilde{u}(y)$, and $\beta \in \Delta(Z)$ be such that $\tilde{u}(\{\beta\}) = 0^*$.

(The existence of such y is guaranteed by regularity condition (c) of ϕ.) Now, let $x = \cup_{s \in S_n}\{\lambda\phi(y,s) + (1-\lambda)\beta\}$. Lemma 2 implies $x \sim \lambda y \oplus (1-\lambda)\{\beta\}$, and thus, we have $J(x) = I(\tilde{u}(x)) = I(\lambda a + (1-\lambda)\tilde{u}(\{\beta\})) = I(\lambda a)$. Next, let β' be such that $\{\beta'\} \sim y$. Then, by Axiom 4 and Lemma 1, $x \sim \lambda y \oplus (1-\lambda)\{\beta\} \sim \lambda\{\beta'\} \oplus (1-\lambda)\{\beta\} \sim \{\lambda\beta' + (1-\lambda)\beta\}$. That is, $J(x) = J(\{\lambda\beta' + (1-\lambda)\beta\}) = \lambda J(\{\beta'\}) + (1-\lambda)J(\{\beta\}) = \lambda I(\tilde{u}(y)) + (1-\lambda)I(\tilde{u}(\{\beta\})) = \lambda I(a)$. Accordingly, we obtain $I(\lambda a) = \lambda I(a)$. Now, define $I(a) = \frac{1}{\lambda}I(\lambda a)$ for all $\lambda > 0$ and $\lambda a \in B_0(K)$. By the positive homogeneity of I on $B_0(K)$ that we have shown, $I(a)$ is homogeneous of degree one for all $a \in B$.

Second, we show that I is C-independent. By homogeneity, 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 a, $\xi^* \in B_0(K)$. Let $x \in \mathcal{A}$ be such that $|x| = n$ and $a = \tilde{u}(x)$, $\beta' \in \Delta(Z)$ be such that $\{\beta'\} \sim x$, and $\beta \in \Delta(Z)$ be such that $\tilde{u}(\{\beta\}) = \xi^*$. By Lemma 2 and Axiom 4, $\cup_{s \in S_n}\{\frac{1}{2}\phi(x,s) + \frac{1}{2}\beta\} \sim \frac{1}{2}x \oplus \frac{1}{2}\{\beta\} \sim \frac{1}{2}\{\beta'\} \oplus \frac{1}{2}\{\beta\} \sim \{\frac{1}{2}\beta' + \frac{1}{2}\beta\}$, which implies that $I(\frac{1}{2}a + \frac{1}{2}\xi^*) = J(\{\frac{1}{2}\beta' + \frac{1}{2}\beta\}) = \frac{1}{2}J(\{\beta'\}) + \frac{1}{2}J(\{\beta\}) = \frac{1}{2}J(x) + \frac{1}{2}J(\{\beta\}) = \frac{1}{2}I(a) + \frac{1}{2}I(\xi^*)$.

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 all a, $b \in B_0(K)$. Let x, $y \in \mathcal{A}$ be such that $|x| = |y| = n$, $a = \tilde{u}(x)$, and $b = \tilde{u}(y)$. Suppose $I(a) = I(b)$, that is, $x \sim y$. Then, it follows from Axiom 5 and Lemma 2 that $x \succsim \frac{1}{2}x \oplus \frac{1}{2}y \sim \cup_{s \in S_n}\{\frac{1}{2}\phi(x,s) + \frac{1}{2}\phi(y,s)\}$, implying 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). Accordingly, 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 completes the proof. The first and third equalities follow from C-independence, while the second inequality follows from $I(a) = I(c)$. $\square$

Lemma 6 *Let I be a monotonic, sublinear, and C-independent functional on B with $I(1^*) = 1$. Then, there is a closed and convex set $\mathcal{M}_n$ of finitely additive probability measures over S_n such that $I(b) = \max_{\mu \in \mathcal{M}_n} \int_{S_n} b d\mu$ for all $b \in B$. Furthermore, $\mathcal{M}_n$ is unique.*

Proof Fix $b \in B$ such that $I(b) > 0$. We first show that there is a (finitely additive) probability measure μ_b over S_n such that $I(b) = \int_{S_n} b d\mu_b$ and $I(a) \geq \int_{S_n} a d\mu_b$ for all $a \in B$. Define

$$D_1 = \mathrm{co}(\{a \in B : a \geq 1^*\} \cup \{a \in B : a \geq b/I(b)\})$$

(where $\mathrm{co}(\cdot)$ denotes the convex hull of the set $(\cdot)$) and

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

Let $d_1 = \lambda a + (1-\lambda)a'$, where $a \geq 1^*$, $a' \geq b/I(b)$, and $\lambda \in [0,1]$. Then, it follows from monotonicity, homogeneity, and C-independence that $I(d_1) \geq \lambda + (1-\lambda)I(a') \geq 1$, which implies that $D_1 \cap D_2 = \emptyset$. Furthermore, both D_1

and D_2 have inner points and are convex (the convexity of D_2 follows from the sublinearity of I). Thus, by a separating hyperplane theorem (e.g., Aliprantis and Border 2006), there is 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 $d_2 \in D_2$. Because 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, because 1^* is a limit point of D_2, the inverse inequality also holds, and we can conclude that $F_b(1^*) = 1$. Furthermore, F_b is nonnegative because, for all nonempty $E \subseteq S_n$ and the indicator function 1_E of E, we have $1^* - 1_E \in D_2$ and $F_b(1_E) + F_b(1^* - 1_E) = F_b(1^*) = 1$, which implies that $F_b(1_E) \geq 0$.

Accordingly, because F_b is a nonnegative linear functional, the Riesz representation theorem (e.g., Aliprantis and Border 2006) implies that there is a finitely additive probability measure μ_b such that $F_b(a) = \int_{S_n} a d\mu_b$ for all $a \in B$. We show that $F_b(a) \leq I(a)$ for all $a \in B$ and $F_b(b) = I(b)$. First, assume that $I(a) > 0$. Because $a/I(a) - (1/m)^* \in D_2$ for all $m \in \mathbb{N}$ 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 $\xi \in \Re$ such that $I(a+\xi^*) > 0$). Second, we focus on the special case $a = b$. Note that $b/I(b) \in D_1$, and so it follows from (6) that $F_b(b) \geq I(b)$. Because the previous argument indicates that the inverse inequality also holds, we have $F_b(b) = I(b)$.

Now, let $\mathcal{M}_n \equiv \overline{\text{co}(\{\mu_b : b \in B, I(b) > 0\})}$, for μ_b defined above. It follows from the previous paragraph that $I(a) \geq \max_{\mu \in \mathcal{M}_n} \int_{S_n} a d\mu$ for all $a \in B$. It has also been shown that, for all $a \in B$ such that $I(a) > 0$, there is a probability measure $\mu_a \in \mathcal{M}_n$ such that $I(a) = \int_{S_n} a d\mu_a$, which implies that $I(a) \leq \max_{\mu \in \mathcal{M}_n} \int_{S_n} a d\mu$. Applying C-independence, a similar argument also holds for $I(a) \leq 0$.

Finally, we show the uniqueness of $\mathcal{M}_n$. Suppose that there are distinct sets $\mathcal{M}_n$ and $\mathcal{M}'_n$ satisfying the statements of this lemma; that is, $I(\tilde{u}(x)) = \max_{\mu \in \mathcal{M}_n} \int_{S_n} u(\phi(x,s)) d\mu(s)$ and $I'(\tilde{u}(x)) = \max_{\mu \in \mathcal{M}'_n} \int_{S_n} u(\phi(x,s)) d\mu(s)$ both represent $\succsim$ for all $x \in \mathcal{A}$ such that $|x| = n$. Choose $\tilde{\mu} \in \mathcal{M}_n \backslash \mathcal{M}'_n$ (if such $\tilde{\mu}$ does not exist, choose $\tilde{\mu} \in \mathcal{M}'_n \backslash \mathcal{M}_n$ instead and proceed accordingly). Then, because $\mathcal{M}'_n$ is convex, the separating hyperplane theorem indicates that there exists $a \in B$ such that $\int_{S_n} a d\tilde{\mu} > \max_{\mu \in \mathcal{M}'_n} \int_{S_n} a d\mu$. It follows from regularity condition (c) that there exists $y \in \mathcal{A}$ such that $|y| = n$ and $a(\cdot) = u \circ \phi(y, \cdot)$, implying $I(\tilde{u}(y)) > I'(\tilde{u}(y))$, which is a contradiction. $\square$

We now conclude the proof of Theorem 1. Because Lemmas 5 and 6 hold for an arbitrary $n \in \mathbb{N}$, Lemmas 1–6 imply that, by defining $S = \{s_1, s_2, \cdots\}$, $\mathcal{M} = \cup_{i=n}^{\infty} \mathcal{M}_n$, and a regular ϕ, we obtain an ASC representation for all $x \in \mathcal{A}$. The essential uniqueness follows from the construction of S, $\mathcal{M}$, and ϕ.

In particular, the uniqueness of ϕ with respect to the relevant mental states is shown as follows. Suppose that there exist regular ϕ and ϕ' representing the preference, and x, $y \in \mathcal{A}$ such that $\phi(x, \tilde{s}) = \phi'(x, \tilde{s})$ and $\phi(y, \tilde{s}) \neq \phi'(y, \tilde{s})$

for some relevant $\tilde{s} \in S$. Without loss of generality, we assume $|x| = |y|$. By construction, $\lambda y \oplus (1-\lambda)x \sim \cup_{s \in S_{|y|}} \{\lambda\phi(y,s)+(1-\lambda)\phi(x,s)\}$ and $\lambda' x \oplus (1-\lambda')y \sim \cup_{s' \in S_{|y|}} \{\lambda'\phi'(x,s') + (1-\lambda')\phi'(y,s')\}$ for all λ, $\lambda' \in [0,1]$, and so Axiom 2b implies that $\lambda'' y \oplus (1-\lambda'')y \sim \{\lambda''\phi(y,s) + (1-\lambda'')\phi'(y,s') : \phi(x,s) = \phi'(x,s')$ for some $s, s' \in S_{|y|}\} \equiv z_{\lambda''}$ for all $\lambda'' \in [0,1]$. However, because $z_{\lambda''}$ includes $\lambda''\phi(y,\tilde{s}) + (1-\lambda'')\phi'(y,\tilde{s})$, $\phi(y,\tilde{s}) \neq \phi'(y,\tilde{s})$, and $\tilde{s}$ is relevant, we conclude without loss of generality that $z_{\lambda''}$ is not indifferent to y for all $\lambda'' \in [0,1]$, contradicting Axiom 2a.

Finally, the desired representation for all random menus P follows from the affineness of u. $\square$

Proof of Theorem 2

The sufficiency ($\mathcal{M}_n = \Delta(S_n)$ implies monotonicity) is straightforward. We show the necessity. First, we prove the following lemma.

Lemma 7 *Suppose that $\succsim$ admits a regular ASC representation $(u, \phi, S, \mathcal{M})$ and satisfies monotonicity. Let $x \in \mathcal{A}$, $\hat{\beta} \in x$, and $\hat{s} \in S_{|x|}$ be such that $\hat{\beta} \succ \beta$ for all $\beta \in x$, $\beta \neq \hat{\beta}$, and $\phi(x,\hat{s}) = \hat{\beta}$. Then, there exists $\mu \in \mathcal{M}_{|x|}$ such that $\mu(\hat{s}) = 1$.*

Proof Let $x \in \mathcal{A}$, $\hat{\beta} \in x$, and $\hat{s} \in S_{|x|}$ be such that $\hat{\beta} \succ \beta$ for all $\beta \in x$, $\beta \neq \hat{\beta}$, and $\phi(x,\hat{s}) = \hat{\beta}$. Suppose that $\mu(\hat{s}) < 1$ for all $\mu \in \mathcal{M}_{|x|}$. Then, we have $u(\hat{\beta}) > V(x) = \max_{\mu \in \mathcal{M}_{|x|}} \int_{S_{|x|}} u(\phi(x,s))d\mu(s)$, or $\{\hat{\beta}\} \succ x$, which contradicts the monotonicity. $\square$

Because Lemma 7 applies to all such x and $\hat{\beta}$, there exists $\mu \in \mathcal{M}_n$ such that $\mu(s) = 1$ for any given $n \in \mathbb{N}$ and $s \in S_n$. Furthermore, because $\mathcal{M}_n$ is closed and convex, $\mathcal{M}_n \supseteq \overline{\mathrm{co}(\{\mu \in \Delta(S_n) : \mu(s) = 1 \text{ for some } s \in S_n\})} = \Delta(S_n)$. $\mathcal{M}_n \subseteq \Delta(S_n)$ is straightforward, which concludes the proof. $\square$

Proof of Theorem 3

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

Conversely, we show that $\succsim_2$ with a greater preference for commitment to a singleton menu than $\succsim_1$ implies $\mathcal{M}^1 \supseteq \mathcal{M}^2$. Assume that there exists $n \in \mathbb{N}$ such that $\mu' \in \mathcal{M}^2_n \setminus \mathcal{M}^1_n$, where $\mathcal{M}^i_n \in \mathcal{M}^i$ for $i = 1$ and 2 denote the sets of probability measures over S_n in each ASC representation. Because $\mathcal{M}^1_n$ is convex, the separating hyperplane theorem implies that there exists $a : S_n \to \Re$ such that $\max_{\mu \in \mathcal{M}^1_n} \int_{S_n} a d\mu < \int_{S_n} a d\mu'$, and so it follows from regularity condition (c) that there exists some $x \in \mathcal{A}$ such that $|x| = n$ and

$a(\cdot) = u \circ \phi(x, \cdot)$. Without loss of generality, we also assume that $\beta \in x$ exists such that $u(\beta) = \int_{S_n} u(\phi(x,s))d\mu'(s) < \max_{\mu \in \mathcal{M}_n^2} \int_{S_n} u(\phi(x,s))d\mu(s)$. Then,

$$\max_{\mu \in \mathcal{M}_n^1} \int_{S_n} u(\phi(x,s))d\mu(s) < \int_{S_n} u(\phi(x,s))d\mu'(s) = u(\beta) < \max_{\mu \in \mathcal{M}_n^2} \int_{S_n} u(\phi(x,s))d\mu(s),$$

implying that $\{\beta\} \succ_1 x$ and $x \succ_2 \{\beta\}$, which is a contradiction. $\square$

Proof of Observation 2

Showing that Axiom 2' implies Axiom 2" is straightforward. Conversely, let x, $y \in \mathcal{A}$ be such that $|x| \leq |y|$. Axiom 2"b (specifically, Definition 1a) implies that there exist a nonempty $C_0 \subseteq x \times y$ and $(\tilde{\alpha}, \tilde{\beta}) \in C_0$ such that $\lambda_0 x \oplus (1-\lambda_0)y \sim \{\lambda_0\alpha + (1-\lambda_0)\beta : (\alpha, \beta) \in C_0\}$ for all $\lambda_0 \in [0,1]$. Next, Axiom 2"a implies that for $C_1 \equiv x \times x$ and all $\lambda_1 \in [0,1]$, $\lambda_1 x \oplus (1-\lambda_1)x \sim \{\lambda_1\alpha + (1-\lambda_1)\alpha' : (\alpha, \alpha') \in C_1\}$, which gives $(\alpha, \tilde{\alpha}) \in C_1$ for all $\alpha \in x$. Axiom 2"a also implies that for $C_2 \equiv y \times y$ and all $\lambda_2 \in [0,1]$, $\lambda_2 y \oplus (1-\lambda_2)y \sim \{\lambda_2\beta' + (1-\lambda_2)\beta : (\beta', \beta) \in C_2\}$, which obtains $(\tilde{\beta}, \beta) \in C_2$ for all $\beta \in y$. The iterative application of Axiom 2"b (specifically, Definition 1b) implies that for $C = \{(\alpha, \beta) : (\alpha, \tilde{\alpha}) \in C_1, (\tilde{\beta}, \beta) \in C_2\} = x \times y$ and all $\lambda \in [0,1]$, $\lambda x \oplus (1-\lambda)y \sim \{\lambda\alpha + (1-\lambda)\beta : (\alpha, \beta) \in C\} = \lambda x + (1-\lambda)y$, which concludes the proof. $\square$

References

Ahn, D.S.: Ambiguity without a state space. *Rev. Econ. Stud.* 75, 3-28 (2007)

Ahn, D.S., Sarver, T.: Preference for flexibility and random choice. *Econometrica* 81, 341-361 (2013)

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

Bandyopadhyay, T., Dasgupta, I., Pattanaik, P.K.: Stochastic revealed preference and the theory of demand. *J. Econ. Theory* 84, 95-110 (1999)

Broadbent, D.E., Cooper, P.F., FitzGerald, P., Parkes, K.R.: The cognitive failures questionnaire (CFQ) and its correlates. *Br. J. Clin. Psychol.* 19, 177-188 (1982)

Bruine de Bruin, W.: Save the last dance for me: unwanted serial position effects in jury evaluations. *Acta Psychol.* 118, 245-260 (2005)

Chandrasekher, M.: Incomplete menu preferences and ambiguity. *mimeo*, Arizona State University.
`http://www.public.asu.edu/~mchandr6/Working%20Papers/incompletemenus.pdf`

(2015). Accessed 15 March 2016

Chatterjee, K., Krishna, R.V.: A “dual self” representation for stochastic temptation. *Am. Econ. J.: Microecon.* 1, 148-167 (2009)

Chatterjee, K., Krishna, R.V.: Uniquely representing “a preference for uniformity.” *B. E. J. Theor. Econ.* 12, 1-34 (2012)

Christenfeld, N.: Choices from identical options. *Psychol. Sci.* 6, 50-55 (1995)

Dekel, E., Lipman, B.: Costly self-control and random self-indulgence. *Econometrica* 80, 1271-1302 (2012)

Dekel, E., Lipman, B., Rustichini, A.: Representing preferences with a unique subjective state space. *Econometrica* 69, 891-934 (2001)

Dekel, E., Lipman, B., Rustichini, A.: Temptation-driven preferences. *Rev. Econ. Stud.* 76, 937-971 (2009)

Epstein, L., Marinacci, M., Seo, K.: Coarse contingencies and ambiguity. *Theor. Econ.* 2, 355-394 (2007)

Ergin, H., Sarver, T.: A unique costly contemplation representation. *Econometrica* 78, 1285-1339 (2010)

Ergin, H., Sarver, T.: Hidden actions and preferences for timing of resolution of uncertainty. *Theor. Econ.* 10, 489-541 (2015)

Fudenberg, D., Levine, D.: A dual-self model of impulse control. *Am. Econ. Rev.* 96, 1449-1476 (2006)

Ghirardato, P., Marinacci, M.: Ambiguity made precise: a comparative foundation. *J. Econ. Theory* 102, 251-289 (2002)

Gilboa, I., Schmeidler, D.: Maxmin expected utility with non-unique prior. *J. Math. Econ.* 18, 141-53 (1989)

Gross, J.J.: The emerging field of emotion regulation: An integrative view. *Rev. Gen. Psychol.* 2, 271-199 (1998)

Gul, F., Pesendorfer, W.: Temptation and self-control. *Econometrica* 69, 1403-1435 (2001)

Gul, F., Pesendorfer, W.: Random expected utility. *Econometrica* 74, 121-146 (2006)

Huber, J., Payne, J., Puto, C.: Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis. *J. Consum. Res.* 9, 90-98 (1982)

Kahneman, D., Wakker, P.P., Sarin, R.: Back to Bentham? Explorations of experienced utility. *Q. J. Econ.* 112, 375-405 (1997)

Kreps, D.M.: A representation theorem for "preference for flexibility." *Econometrica* 47, 565-577 (1979)

Kreps, D.M.: Static Choice in the Presence of Unforseen Contingencies, In: Dasguputa, P., Gale, D., Hart, O., Maskin, E. (eds.), Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn, MIT Press, Cambridge, MA., 258-281 (1992).

Krueger, L.E., Salthouse, T.A.: Influence of cognitive abilities and age on word recall performance across trials and list segments. *Am. J. Psychol.* 124, 291-300 (2011)

Larsen, R.J.: Toward a science of mood regulation. *Psychol. Inq.* 11, 129-141 (2000)

Loewenstein, G., O'Donoghue, T.: Animal spirits: affective and deliberative processes in economic behavior. CAE Working Paper 04-14, Cornell University. `https://cae.economics.cornell.edu/04-14.pdf` (2004). Accessed 15 March 2016

Luce, R.D.: *Individual Choice Behavior*. John Wiley & Sons, New York, NY (1959)

Luce, R.D., Suppes, P.: Preference, utility and subjective probability. In: Luce, R.D., Bush, R.R., Galanter, E. (eds.), *Handbook of Mathematical Psychology*, Vol III, pp. 249-410. Wiley, New York, NY (1965)

Manzini, P., Mariotti, M.: Stochastic choice and consideration sets. *Econometrica* 82, 1153-1176 (2014)

Manzini, P., Mariotti, M.: Imperfect attention and menu evaluation. *mimeo*, University of St. Andrews.
`http://www.st-andrews.ac.uk/~wwwecon/repecfiles/4/1319.pdf` (2015) Accessed 15 March 2016

Masatlioglu, Y., Nakajima, D., Ozbay, E.Y.: Revealed attention. *Am. Econ. Rev.* 102, 2183-2205 (2012)

Murdock, B.B.: The serial position effect of free recall. *J. Exp. Psychol.* 64, 482-488 (1962)

Nehring, K.: Preference for flexibility in a Savage framework. *Econometrica* 67, 101-119 (1999)

Nehring, K.: Self-control through second-order preferences. *mimeo*, University of California, Davis.
`http://www.parisschoolofeconomics.eu/IMG/pdf/sop-9-06.pdf` (2006) Accessed 15 March 2016

Nishimura, K.G., Ozaki, H.: An axiomatic approach to epsilon-contamination. *Econ. Theory* 27, 333-340 (2006)

Noor, J., Takeoka, N.: Uphill self-control. *Theor. Econ.* 5, 127-158 (2010)

Noor, J., Takeoka, N.: Menu-dependent self-control. *J. Math. Econ.* 61, 1-20 (2015).

Olszewski, W.: Preferences over sets of lotteries. *Rev. Econ. Stud.* 74, 567-595 (2007)

Piccione, M., Rubinstein, A.: On the interpretation of decision problems with imperfect recall. *Games Econ. Behav.* 20, 3-24 (1997)

Posner, M.I., Snyder, C.R.R.: Attention and cognitive control. In: Solso, R.L. (ed.), *Information Processing and Cognition: The Loyola Symposium*, pp. 55-85. Lawrence Erlbaum Associates, Hillsdale, NJ (1975)

Rubinstein, A., Salant, Y.: A model of choice from lists. *Theor. Econ.* 1, 3-17 (2006)

Sarver, T.: Anticipating regret: why fewer options may be better. *Econometrica* 76, 263-305 (2008)

Schmeidler, D.: Subjective probability and expected utility without additivity. *Econometrica* 57, 571-587 (1989)

Selten, R.: Reexamination of the perfectness concept for equilibrium points in extensive games. *Int. J. Game Theory* 4, 25-55 (1975)

Sherman, J.W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T.J., Groom, C.J.: The self-regulation of automatic associations and behavioral impulses. *Psychol. Rev.* 115, 314-335 (2008)

Simonson, I.: Choice based on reasons: The case of attraction and compromise effects. *J. Consum. Res.* 16, 158-174 (1989)

Slovic, P.: Choice between equally-valued alternatives. *J. Exp. Psychol.* 1, 280-287 (1975)

Strotz, R.H.: Myopia and inconsistency in dynamic utility maximization. *Rev. Econ. Stud.* 23, 165-180 (1955)

Tentori, K., Osherson, D., Hasher, L., May, C.: Wisdom and aging: Irrational preferences in college students but not older adults. *Cognition* 81, B87-B96 (2001)

Thaler, R.H., Shefrin, H.M.: An economic theory of self-control. *J. Polit. Econ.* 89, 392-406 (1981)

Tyszka, T.: Contextual multiattribute decision rules. In: Sjoberg, I., Tyszka, T., Wise, J. (eds.), *Human Decision Making*, pp. 243-256. Doxa, Bodafors, Sweden (1983)

Wright, P., Barbour, F.: Phased decision strategies: sequels to an initial screening. In: Starr, M., Zeleny, M. (eds.), *Studies in Management Sciences, Multiple Criteria Decision Making*, pp. 91-109. North-Holland, Amsterdam, Netherlands (1977)