Deliberately Stochastic*

Simone Cerreia-Vioglio[†], David Dillenberger[‡], Pietro Ortoleva[§], Gil Riella[¶]

First version: February 2012
This version: May 5, 2018

Abstract
We study stochastic choice as the outcome of *deliberate* randomization. We derive a general representation of a stochastic choice function where stochasticity allows the agent to achieve from any set the maximal element according to underlying preferences over lotteries. In this model, stochasticity in choice captures complementarity between elements in the set, and thus necessarily implies violations of Regularity/Monotonicity, one of the most common properties of stochastic choice. This feature separates our approach from other models (e.g., Random Utility). We also characterize the case where preferences follow Cautious Expected Utility. Here, stochastic choice is linked to Certainty Bias.

JEL: *D80, D81*

Keywords: *Stochastic Choice, Random Utility, Hedging, Cautious Expected Utility, Convex Preferences, Regularity, Certainty Bias.*

*We thank Eddie Dekel, Kfir Eliaz, Mira Frick, Faruk Gul, Ryota Iijima, Peter Klibanoff, Jay Lu, Efe Ok, Wolfgang Pesendorfer, Marciano Siniscalchi, Tomasz Strzalecki, and the audience at various seminars and conferences for their useful comments. Cerreia-Vioglio gratefully acknowledges the financial support of ERC grant SDDM-TEA, Ortoleva gratefully acknowledges the financial support of NSF grant SES-1559462, and Riella gratefully acknowledges the financial support of CNPq of Brazil, grant no. 304560/2015-4. Part of this work was done while Dillenberger was visiting the Economics department at Princeton University. He is most grateful to this institution for its hospitality.
[†]Department of Decision Sciences, Università Bocconi, email: `simone.cerreia@unibocconi.it`.
[‡]Department of Economics, University of Pennsylvania, email: `ddill@sas.upenn.edu`.
[§]Department of Economics and Woodrow Wilson School, Princeton University, email: `pietro.ortoleva@princeton.edu`.
[¶]FGV EPPG Escola de Políticas Públicas e Governo, email: `gil.riella@fgv.br`.

1 Introduction

A robust finding in the study of individual decision-making is the presence of stochastic, or random, choice: when subjects are asked to choose from the same set of options multiple times, they often make different choices.[1] An extensive literature has documented this pattern in many experiments, in different settings and with different populations, both in the lab and in the field. It often involves a significant fraction of the choices, even when subjects have no value for experimentation (e.g., when there is no feedback), or when there are no bundle or portfolio effects (e.g., when only one choice is paid).[2] It thus appears incompatible with the typical assumption in economics that subjects have a complete and stable preference ranking over the available alternatives and consistently choose the option that maximizes it.

A large body of theoretical work has developed models to capture stochastic behavior. Most of these models can be ascribed to one of two classes. First, models of "Random Utility/Preferences," according to which subjects' answers change because their preferences change stochastically.[3] Second, models of "bounded rationality," or "mistakes," according to which subjects have stable and complete preferences, but may fail to always choose the best option and thus exhibit a stochastic pattern.[4]

While according to the interpretations above the stochasticity of choice happens involuntarily, a third possible interpretation is that stochastic choice is a *deliberate* decision of the agent: she may *choose* to report different answers. The goals of this paper are: 1) to develop axiomatically models in which stochastic choice follows this interpretation; 2) to identify whether, and how, such models of deliberate randomization generate different behaviors and can be distinguished from existing models of stochastic choice; and 3) to show how stochastic choice can be seen as stemming from the same source as known violations of Expected Utility, like the Certainty Bias.

A small existing literature has suggested why subjects may wish to report stochastic answers. Machina (1985) notes that this is precisely what the agent may wish to do

[1]To avoid confusion, these terms are used to denote two different phenomena: 1) *one* person faces the same question *multiple times* and gives different answers; 2) *different* subjects answer the same question only *once*, but subjects who appear similar, given the available data, make different choices. In this paper we focus on the first one.

[2]The pattern of stochastic choice was first reported in Tversky (1969). A large literature followed: focusing on choices between risky gambles (as in our model), see Camerer (1989), Starmer and Sugden (1989), Hey and Orme (1994), Ballinger and Wilcox (1997), Hey (2001), Regenwetter et al. (2011), Regenwetter and Davis-Stober (2012), and Agranov and Ortoleva (2017).

[3]Thurstone (1927), Luce (1959), Harsanyi (1973), Falmagne (1978), Cohen (1980), Barberá and Pattanaik (1986), McFadden and Richter (1991), Loomes and Sugden (1995), Clark (1996), McFadden (2006), Gul and Pesendorfer (2006), Ahn and Sarver (2013), Fudenberg and Strzalecki (2015).

[4]Models of this kind appear in economics, psychology and neuroscience, including the well-known Drift Diffusion model: among many, Busemeyer and Townsend (1993), Harless and Camerer (1994), Hey and Orme (1994), Camerer and Ho (1994), Wu and Gonzalez (1996), Ratcliff and McKoon (2008), Gul et al. (2014), Manzini and Mariotti (2014), Woodford (2014), Fudenberg and Strzalecki (2015). For surveys, Ratcliff and Smith (2004), Bogacz et al. (2006), Johnson and Ratcliff (2013).

if her preferences over lotteries or acts are *convex*, that is, exhibit affinity towards randomization between equally good options. Crucially, convexity is a property shared by many existing models of decision making under risk, and it captures ambiguity aversion in the context of decision making under uncertainty. Convexity of preferences also has experimental support (Becker et al., 1963; Sopher and Narramore, 2000). Different reasons for stochastic choice to be deliberate were suggested by Marley (1997) and Swait and Marley (2013), who follow lines similar to Machina (1985); Dwenger et al. (2016), that suggest it may be due to a desire to minimize regret; and Fudenberg et al. (2015), who connect it to uncertain taste shocks. In Section 5 we discuss these papers in detail.

Recent experimental evidence supports the interpretation of stochastic choice as deliberate. Agranov and Ortoleva (2017) show how subjects give different answers also when the same question is asked three times in a row and subjects are aware of the repetition; they seem to explicitly choose to report different answers.[5] Dwenger et al. (2016) find that a large fraction of subjects choose lotteries between available allocations, indicating an explicit preference for randomization. They also show similar patterns using the data from a clearinghouse for university admissions in Germany, where students must submit multiple rankings of the universities they would like to attend. These are submitted at the same time, but only one of them (chosen randomly) matters. They find that a significant fraction of students report inconsistent rankings, even when there are no strategic reasons to do so. A survey among applicants supports the interpretation that these random allocations are chosen intentionally, and show that they are correlated with an explicit preference for randomization.[6]

We develop axiomatically models of stochastic choice over lotteries as the outcome of a deliberate desire to report a stochastic answer. We aim to capture and formalize the intuition of Machina (1985) that such inclination may be a rational reaction if the underlying preferences over lotteries are (at least locally) convex. We consider a stochastic choice function over sets of lotteries over monetary outcomes, which assigns to any set a probability distribution over its elements. We focus on lotteries not for technical reasons, but because we are interested in linking stochastic choice to features of preferences over lotteries in general, and violations of Expected Utility in particular. The presence of lotteries is thus essential to make the connection. The focus on lotteries over *monetary* outcomes, however, is *inessential* to the results in the first part of the paper (Sections 2 and 3), as similar results are available for arbitrary prize spaces (see footnote 10 below). The specific model we consider in Section 4 uses certainty equivalents and thus requires monetary lotteries. To avoid confusion, we will use the same domain throughout the paper.

[5] In a survey conducted at the end of the experiment, most subjects report choosing different answers deliberately. These results hold true also in robustness tests with unusually high stakes.

[6] Kircher et al. (2013) consider a version of the dictator game in which dictators can choose between 7.5 euros for themselves and 0 to the recipient, 5 to both, or a lottery between them. About one third of the subjects chooses to randomize.

We begin our analysis with a general representation theorem: we show that a rationality-type condition on stochastic choice, reminiscent of known acyclicity conditions, guarantees that it can be represented as if the agent had a preference relation over the final monetary lotteries, and chose the *optimal* mixing over the existing options. We derive our result showing a parallel between this approach and the literature on choice from limited datasets – since assuming that the agent considers all possible randomizations is equivalent to saying that choice is observable only from convex sets. In this model the stochasticity has a purely *instrumental* value for the agent: she does not value the randomization *per se*, but rather because it allows her to obtain the lottery over final outcomes she prefers. Implicit in this approach is that agents evaluate mixtures of lotteries by looking at the distribution over final outcomes they induce.

Next, we show that the general model above has some stark implications. Possibly the most well-known property of stochastic choice, widely used in the literature, is Regularity (also called Monotonicity): it posits that the probability of choosing p from a set cannot decrease if we remove elements from it. It is often seen as the stochastic equivalent of independence of irrelevant alternatives (IIA), and it is satisfied by many models in the literature – most prominently, models of Random Utility, albeit it is well-known that it is often empirically violated. We show that our model of deliberate stochastic choice *must* violate Regularity (unless the stochastic choice is degenerate, i.e., there is no stochasticity). Intuitively, our agent may choose from a set A two options that, together, allow her to "hedge." But this holds only if they are *both* chosen: they are complementary to each other. If either option is removed from A, the possibility of hedging disappears and the agent no longer has incentive to pick the remaining one. This generates a violation of Regularity. The key observation is that the agent considers all the elements chosen *as a whole*, for the general hedging they provide *together*. By contrast, Regularity is based on the assumption that the appeal of each option is independent from the other options present in the menu or in the choice. Thus, a violation of Regularity is an essential feature of the hedging behavior that we aim to capture – as we formally show.

That the model above is inconsistent with Regularity has direct implications on its relation with existing models, most prominently models of Random Utility. Since it is well-known that the latter must satisfy Regularity, the only behavior that can be represented by both models is one that can always be described as if the agent had only one utility and randomization occurs only in the case of indifference – a degenerate random utility. Thus, the conceptual difference between the two models is reflected in a substantial behavioral difference, via the property of Regularity, which is very easily testable in experiments. Since also the models in Fudenberg et al. (2015) satisfy Regularity, the same relation holds between our model and theirs.

Our last set of results derives a special case of our general representation. Since a desire to mix must derive from violations of Expected Utility, additional structure will result from imposing consistency on when such violations can occur. Since one of the most robust findings of such violations is the Certainty Bias, as captured by both Allais paradoxes (Common-Ratio and Common-Consequence effects), we posit that violations cannot occur in ways that are explicitly in the *opposite* direction, at least in the extreme case in which the stochastic choice function is degenerate.[7] We show that this postulate (with continuity and risk aversion), characterizes the special case of the model above in which the underlying preferences are represented by the Cautious Expected Utility model of Cerreia-Vioglio et al. (2015): the agent has a *set* of utility functions over outcomes, and evaluates each lottery p by first computing the certainty equivalent of p with respect to each possible function in the set, and then picks the smallest one. These preferences are convex, and thus display (weak) preference for mixing. In this model there are multiple utilities being considered, which is similar to what happens with models of Random Utility, where one utility is randomly picked each time a choice is made. Here, instead, all utilities are considered *at the same time*, and the agent uses the one that returns the lowest certainty equivalent. It is as if the agent were aware – or *meta-cognitive* – of the presence of multiple utilities and acted with caution given this awareness. And indeed, as we have already seen, our model is behaviorally distinct from Random Utility.

The Cautious Stochastic Choice model tightly links stochasticity of choice and violations of Expected Utility as the Certainty Bias. Both stem from the presence of multiple utilities and the use of caution. This connection is formal. We show that as long as there are finitely many utilities, agents have a strict desire to randomize if and only if they violate Expected Utility in line with the Certainty Bias, which in turn holds if and only if her set of utilities contains more than one element. Our model thus belongs to the small literature that suggests some unification of different deviations from the 'standard' model, linking them to the same source: here stochastic choice and Certainty Bias are both due to the presence of multiple utilities and caution.[8]

The remainder of the paper is organized as follows. Section 2 presents the general Deliberate Stochastic Choice model. Section 3 establishes that our model is incompatible with Regularity. Section 4 presents the special case of the Cautious Stochastic Choice model and Section 5 discusses the relation with existing literature. All proofs appear in the Appendix.

---

[7]Our axiom will be reminiscent of Negative Certainty Independence of Dillenberger (2010). But while the latter is imposed on preferences, here preferences are not observable. Our axiom will be instead imposed *only* in the extreme situations in which the stochastic choice function is degenerate; it posits no restrictions when the stochastic choice function is not degenerate. It is thus conceptually much weaker.

[8]See Dillenberger (2010), Ortoleva (2010), Andreoni and Sprenger (2010), Epper and Fehr-Duda (2015), Dean and Ortoleva (2015), Dean and Ortoleva (2017).

2 A General Model of Deliberately Stochastic Choice

2.1 Framework and Foundations

Let $[w, b] \subset \mathbb{R}$ be a non-trivial interval of monetary prizes and let Δ be the set of lotteries (Borel probability measures) over $[w, b]$, endowed with the topology of weak convergence. We use x, y, z and p, q, r for generic elements of $[w, b]$ and Δ, respectively. Denote by $\delta_x \in \Delta$ the degenerate lottery (Dirac measure at x) that gives the prize $x \in [w, b]$ with certainty. If p and q are such that p strictly first order stochastically dominates q, we write $p >_{FOSD} q$.

Denote by $\mathcal{A}$ the collection of all finite and nonempty subsets of Δ. For any $A \in \mathcal{A}$, co(A) denotes the convex hull of A, that is, $\text{co}(A) = \{\sum_j \alpha_j p_j : p_j \in A \text{ and } \alpha_j \in [0, 1], \sum_j \alpha_j = 1\}$.

The primitive of our analysis is a *stochastic choice function* ρ over $\mathcal{A}$, i.e., a map ρ that associates to each $A \in \mathcal{A}$ a probability measure $\rho(A)$ over A. For any stochastic choice function ρ, $A \in \mathcal{A}$, and $p \in A$, $\text{supp}_\rho(A)$ denotes the support of $\rho(A)$, and we write $\rho(A)(p)$ to denote the probability ρ assigns to p in menu A.

As a final bit of notation, since $\rho(A)$ is a probability distribution over lotteries, thus a compound lottery, we can compute the induced lottery over final monetary outcomes. Denote it by $\overline{\rho(A)} \in \Delta$, that is,

$$\overline{\rho(A)} = \sum_{q \in A} \rho(A)(q)q.$$

By construction, the convex hull of a set A, $co(A)$, will also correspond to the set of all monetary lotteries that can be obtained by choosing a specific ρ and computing the distribution over final prizes it induces.[9]

We can now discuss our first axiom. Our goal is to capture behaviorally an agent who is *deliberately* choosing her stochastic choice function following an underlying preference relation over lotteries. When asked to choose from a set A, she considers all lotteries that can be obtained from A by randomizing: using our notation above, she considers the whole $co(A)$, and the lottery $\overline{\rho(A)}$ can be seen as her 'choice.'

Our axiom is a rationality-type postulate for this case. Consider two sets A_1 and A_2, and suppose that $\overline{\rho(A_2)} \in co(A_1)$. This means that the lottery chosen from A_2 could be obtained also from A_1. Standard rationality posits that the 'choice' from A_1, $\overline{\rho(A_1)}$, must then be at least as good as anything that can be obtained from A_2. Since we do not observe the preferences, we cannot impose this; but at the very least we can say that there cannot be anything in A_2 that strictly first order stochastically dominates $\overline{\rho(A_1)}$. This is the content of our axiom, extended to any sequence of length k of sets.

[9] That is, by construction we must have $co(A) = \{p \in \Delta : p = \overline{\rho(A)}$ for some stochastic choice function $\rho\}$.

Axiom 1 (Rational Mixing)**.** *For each $k \in \mathbb{N}\backslash\{1\}$ and $A_1, \ldots, A_k \in \mathcal{A}$, if*

$$\overline{\rho(A_2)} \in co(A_1), \ldots, \overline{\rho(A_k)} \in co(A_{k-1}),$$

then $q \in co(A_k)$ implies $q \not>_{FOSD} \overline{\rho(A_1)}$.

Rational Mixing is related to conditions of rationality and acyclicity typical in the literature on revealed preferences with limited observations, along the lines of Afriat's condition and the Strong Axiom of Revealed Preferences (see, e.g., Chambers and Echenique 2016). Intuitively, the ability to randomize allows the agent to choose any option in the convex hull of all sets; thus, it is as if we could only see the choices from convex sets, and posit a rationality condition for this case.

Note that Rational Mixing implicitly 1) includes a form of coherence with strict first order stochastic dominance, and 2) assumes that the agent cares only about the *induced distribution over final outcomes*, rather than the *procedure* in which it is obtained. That is, for the agent the stochasticity is *instrumental* to obtain a better distribution over final outcomes, rather than being valuable *per se.* This implies a form of reduction of compound lotteries, which we will maintain throughout.

2.2 Deliberate Stochastic Choice Model

Definition 1. A stochastic choice function ρ admits a Deliberate Stochastic Choice representation if there exists a complete preorder (a transitive and reflexive binary relation) $\succsim$ over Δ such that:

1. For every $A \in \mathcal{A}$
$$\overline{\rho(A)} \succsim q \quad \text{for every } q \in co(A);$$
2. For every pair $p, q \in \Delta$, $p >_{FOSD} q$ implies $p \succ q$.

Theorem 1. *A stochastic choice function ρ satisfies Rational Mixing if and only if it admits a Deliberate Stochastic Choice representation.*

A Deliberate Stochastic Choice model captures a decision maker who has preferences $\succsim$ over lotteries and chooses deliberately the randomization that generates the optimal mixture among existing options. This is most prominent when $\succsim$ is convex and, in particular, if there exist some $p, q \in \Delta$ and $\alpha \in (0, 1)$ such that $\alpha p + (1 - \alpha)q \succ p, q$. When faced with the choice from $\{p, q\}$, she would strictly prefer to randomly choose rather than to pick either of the two options. The stochasticity is thus an expression of the agent's preferences.[10]

[10]As we pointed out in the Introduction, Theorem 1 does not require the lotteries to be over monetary outcomes; they could be over an arbitrary set of prizes, as long as there is some natural dominance relation (a partial order) $\rhd$ over the space of lotteries that one can use to replace $>_{FOSD}$ in the statement of the Rational Mixing axiom. This can be a generalization of the concept of first order stochastic dominance, or it can be any other partial order that satisfies the additional property that $p \rhd q$ implies that $p \rhd \alpha p + (1 - \alpha)q$ for every $\alpha \in [0, 1)$ and for every pair of lotteries p and q.

The Deliberate Stochastic Choice model is very general and does not restrict preferences to be convex. It permits desire for randomization, in regions where strict convexity holds; indifference to randomization, e.g., when $\succsim$ follows Expected Utility, or satisfies Betweenness;[11] or even aversion to randomization, e.g., if $\succsim$ are Rank Dependent Expected Utility (RDU) preferences with pessimistic distortions: in these cases the agent has no desire to mix and we should observe no stochasticity (except for indifferences).[12]

Note also that the Deliberate Stochastic Choice model puts no restriction on the way the agent resolves indifferences: when multiple alternatives maximize the preference relation, any could be chosen. Although it is a typical approach not to rule how indifferences are resolved, this may however lead to discontinuities.[13] This implies that, for each preference $\succsim$, there is more than one stochastic choice function derived from it – depending on how indifferences are resolved. In general, we say that a stochastic choice function $\hat{\rho}$ is consistent with a preference $\succsim$ if and only if for each $A \in \mathcal{A}$, $\overline{\hat{\rho}(A)} \succsim q$ for all $q \in co(A)$.

Remark 1. Our framework implicitly assumes that we observe the stochastic choice function ρ for all sets in $\mathcal{A}$. This is very demanding, and a natural question is what tests are required if we observe only limited data. In fact, the Rational Mixing axiom is necessary and sufficient in *any* dataset that includes all doubletons $\{p, q\}$ such that $p >_{FOSD} q$. Consider any $\mathcal{B} \subset \mathcal{A}$ such that $\{p, q\} \in \mathcal{B}$ whenever $p >_{FOSD} q$ and denote by $\rho_{\mathcal{B}}$ the restriction of ρ on $\mathcal{B}$. Then, $\rho_{\mathcal{B}}$ satisfies Rational Mixing if and only if $\rho_{\mathcal{B}}$ admits a Deliberate Stochastic Choice representation. (The proof follows exactly the same steps as the proof of Theorem 1.)

Remark 2. The preference relation $\succsim$ in a Deliberate Stochastic Choice model need not admit a utility representation. As usual, for this to be guaranteed we need $\succsim$ to be continuous. We call this case a *Continuous* Deliberate Stochastic Choice model. Proposition 3 in Appendix B gives an axiomatic characterization of it, obtained by strengthening Rational Mixing with a continuity requirement. In words, consider the binary relation R, defined as pRq if $p = \overline{\rho(A)}$ for some A such that $q \in co(A)$. Rational Mixing simply posits that the transitive closure of R is consistent with $>_{FOSD}$. To obtain a continuous representation, we need to extend this requirement to the closed transitive closure of R (i.e., its continuous extension).

[11] That is, $p \sim q \Rightarrow \alpha p + (1-\alpha)q \sim q$ for all $p, q \in \Delta$, $\alpha \in (0,1)$. See Dekel (1986); Chew (1989).

[12] If we order the prizes in the support of a finite lottery p, with $x_1 < x_2 < ... < x_n$, then the functional form for RDU is: $V(p) = u(x_n) f(p(x_n)) + \sum_{i=1}^{n-1} u(x_i)[f(\sum_{j=i}^{n} p(x_j)) - f(\sum_{j=i+1}^{n} p(x_j))]$, where $f : [0,1] \to [0,1]$ is strictly increasing and onto and $u : [w, b] \to \mathbb{R}$ is increasing. We say that distortions are pessimistic if f is convex, which implies *aversion* to randomization.

[13] While with choice correspondences the continuity of the underlying preference relation implies continuity of the choice correspondence (i.e., satisfies the closed graph property), here it is as if we observed also the outcome of how indifference is resolved (which may be stochastic). This will necessarily imply discontinuities of ρ, following standard arguments. An alternative, although significantly less appealing, approach would be to consider a stochastic choice *correspondence*, which could be fully continuous.

3 Regularity and Deliberate Randomization

In this section we study the relation of the Deliberate Stochastic Choice model with a well-known property of stochastic choice, extensively used in the literature: Regularity, also called Monotonicity.

Axiom 2 (Regularity)**.** *For each $A, B \in \mathcal{A}$ and $p \in A$, if $A \subseteq B$, then $\rho(B)(p) \leq \rho(A)(p)$.*

Intuitively, Regularity states that if we remove some elements from a set, the probability of choosing the remaining elements cannot decrease. Conceptually, it is related to notions of independence of irrelevant alternatives applied to a stochastic setting: the removal of any element, chosen or unchosen, cannot 'hurt' the chances of choosing any of the remaining ones. In other words, the attractiveness of an option should not depend on the availability of other ones. Crucially, the property of Regularity is one of the characterizing features of models of Random Utility.

To analyze this property in our model, it will be useful to formally define when is the individual exhibiting stochastic choice. Recall that in our model this may happen either when there is a genuine desire to randomize, or in the case of indifferences. We say that an agent exhibits a non-degenerate stochastic choice function if stochasticity is present beyond indifference: if we can find some p and q such that the agent randomizes between them and also when either is made a "little bit worse" by mixing with δ_w (the worst possible outcome).

Definition 2. A stochastic choice function ρ is *non-degenerate* if there exist $p, q \in \Delta$ with $|\text{supp}_\rho(\{p, q\})| \neq 1$ and $\lambda \in (0, 1)$ such that

$$|\text{supp}_\rho(\{\lambda p + (1-\lambda)\delta_w, q\})| \neq 1 \quad \text{and} \quad |\text{supp}_\rho(\{p, \lambda q + (1-\lambda)\delta_w\})| \neq 1.$$

Endowed with these definitions, we have the following theorem.

Theorem 2. *Let ρ be a stochastic choice function that admits a continuous Deliberate Stochastic Choice representation $\succsim$. The following statements are equivalent:*

(i) ρ is non-degenerate;

(ii) ρ and any other $\hat{\rho}$ consistent with $\succsim$ violates Regularity.

(iii) $\succsim$ has a point of strict convexity, that is, there exist $p, q \in \Delta$ and $\lambda \in (0, 1)$ such that

$$\lambda p + (1-\lambda)\, q \succ p, q.$$

Theorem 2 shows that the Deliberate Stochastic Choice model *must* lead to violations of Regularity, unless the Stochastic Choice is degenerate (no stochasticity, or purely to break indifferences). In fact, the result is stronger: violations of Regularity and stochasticity imply one another; and both occur if and only if the underlying preferences have points of strict convexity: without them, under the model there should never be any stochasticity, as the agent does not have a desire to randomize; but with them, we also have violations of Regularity.

One important implication of the result above is to distinguish our model from models that satisfy Regularity, which include most popular models, such as the models of Random Utility and Luce (1958)'s model. Thus, the desire to randomize not only is different, but also leads to a behavior that violates the core property of many models in the literature.[14]

To gain an intuition, consider some p, q where preferences are strictly convex, i.e., there exists $\lambda \in (0,1)$ such that $\lambda p + (1-\lambda) q \succ p, q$, as in item (iii) of the theorem. For simplicity, suppose that $r = \hat{\lambda} p + (1-\hat{\lambda}) q$ is the unique $\succsim$-optimal mix between p and q. Let r_ε be a lottery within distance $\varepsilon > 0$ of r but FOSD dominated by it. First observe that p will be chosen with probability $\hat{\lambda}$ from $\{p, q, r_\varepsilon\}$: in the face of both p and q, the presence of r_ε is of no value to the agent. But suppose q is removed: then, as long as ε is small enough (r_ε is close to r), p will be chosen with very small probability from $\{p, r_\varepsilon\}$: the value of p decreases significantly without q, and the agent now puts most weight on r_ε. This pair of choices violates Regularity. The idea behind this construction is that p and q are *complementary* to each other. But if q is removed, the agent can no longer hedge between them; r_ε, which was inferior to their mixture then becomes an attractive alternative in the face of p. Overall, the crucial aspect is that the ability of choosing *both* p and q at the same time renders them appealing, while they would not be appealing in isolation. This is a fundamental aspect of when hedging is advantageous: the *whole* set of chosen elements is relevant for the agent, for the hedging opportunities it provides. Such complementarity between alternatives violates standard independence of irrelevant alternatives arguments, according to which chosen elements should be appealing in isolation, which is also reflected in the Regularity axiom. For that reason, violations of Regularity are a "structural" feature of our model.[15]

4 A Special Case: Cautious Stochastic Choice

The model characterized in Theorem 1 has the benefits and drawbacks of generality: it captures stochastic choice as the deliberate desire to report a stochastic answer,

[14]We should stress that in this case we do not view violations of Regularity as mistakes or as forms of bounded rationality. Indeed, our model entails a strong form of "rationality:" the agent acts as if she foresees the consequences of each possible randomization and chooses the best one. That her behavior violates Regularity is simply a manifestation of her underlying well-defined preferences over lotteries that have a point of strict convexity.

[15]Our model would in general satisfy a weaker version that posits that choice probabilities do not decrease if we remove elements that are *never chosen* – violations of this may occur because of indifferences.

with few further assumptions; but it puts only minimal restrictions on preferences over lotteries. We now turn to a special case in which we give a specific functional form representation to the underlying preferences $\succsim$.

Recall that the agent may strictly prefer to mix only if the underlying preferences violate Expected Utility – otherwise no mixing is beneficial. To gain more structure, we can restrict how these violations may occur. One of the most robustly documented instances of violation of Expected Utility is the so-called Certainty Effect, as captured, for example, by Allais' Common Ratio and Common Consequences effects: intuitively, agents violate the Independence axiom by over-valuing degenerate lotteries. In desiring to restrict violations of Expected Utility, it is thus natural to posit that such violations cannot be in the unequivocally *opposite* direction.

Suppose that $\{p\} = \text{supp}_\rho(\{p, \delta_x\})$: from the set $\{p, \delta_x\}$, the agent *always* chooses p uniquely. This means that p is more attractive than δ_x, even though the latter is a degenerate lottery and thus potentially very attractive for an agent who may be certainty biased. Suppose now that we mix both options with a lottery q, obtaining $\{\lambda p + (1 - \lambda)q, \lambda\delta_x + (1 - \lambda)q\}$. By doing so we are transforming the sure amount into a lottery. And if p was appealing against it before, the mixture of p should be all the more appealing now that the alternative is no longer certain. And if we replace δ_x with δ_y for some $y < x$ in the mixture, then this should be even more true. This leads us to the following axiom.

Axiom 3 (Weak Stochastic Certainty Effect)**.** *For each $p, q \in \Delta$, $x, y \in [w, b]$ with $x > y$, and $\lambda \in (0, 1]$, if*

$$\{p\} = \text{supp}_\rho(\{p, \delta_x\}),$$

then

$$\{\lambda p + (1 - \lambda)q\} = \text{supp}_\rho(\{\lambda p + (1 - \lambda)q, \lambda\delta_y + (1 - \lambda)q\}).$$

While the axiom above would be satisfied by any agent with Expected Utility preferences, it would also be satisfied by an agent who is certainty biased (see, for example, Kahneman and Tversky 1979). What the axiom rules out are agents who are strictly certainty "averse," as they may choose p uniquely when the alternative is δ_x, but may pick a mixture of the latter if it is no longer degenerate. For example, the axiom rules out violations of Expected Utility that are the opposite of those in Allais' paradoxes.

Weak Stochastic Certainty Effect is related to Negative Certainty Independence, or NCI (Dillenberger, 2010; Cerreia-Vioglio et al., 2015), which is imposed on preferences over lotteries. NCI requires that if p is preferred to δ_x, then a mixture of p and q should be weakly preferred to a mixture (with the same proportions) of δ_x and q. Both axioms follow a similar logic in ruling out the opposite of Certainty Bias. Where they differ, however, is that NCI is imposed on preferences, which here we cannot observe. Weak Stochastic Certainty Effect is instead imposed on a stochastic choice function and, crucially, only in the extreme case in which the stochastic choice function is *degenerate*. It imposes no restrictions on behavior for sets where the stochastic choice function is not degenerate, and it holds vacuously if ρ is never degenerate. It is thus conceptually much weaker than imposing NCI on the underlying preferences – even if there was a way to do so.

We are left with two standard postulates: Continuity and Risk Aversion. Since we posit no restrictions on what the agent does in the case of indifferences, we have to allow for discontinuities of ρ due to this, as we have discussed above (see footnote 13).

Axiom 4 (Continuity)**.** *Let $(p^m) \in \Delta^\infty$ and $(x^m) \in [w,b]^\infty$ be convergent sequences with $p^m \to p$ and $x^m \to x$. Let $y \in (w,x)$ and $q \in \Delta$ be such that $q >_{FOSD} p$. Then:*

- *$\{p^m\} = \mathrm{supp}_\rho(\{p^m, \delta_x\})$ for every m implies $\{p\} = \mathrm{supp}_\rho(\{p, \delta_y\})$. Similarly, $\delta_y \in \mathrm{supp}_\rho(\{p^m, \delta_y\})$ for every m implies $\{\delta_x\} = \mathrm{supp}_\rho(\{p, \delta_x\})$;*
- *$\{p\} = \mathrm{supp}_\rho(\{p, \delta_{x^m}\})$ for every m implies $\{q\} = \mathrm{supp}_\rho(\{q, \delta_x\})$. Similarly, $\delta_{x^m} \in \mathrm{supp}_\rho(\{q, \delta_{x^m}\})$ for every m implies $\{\delta_x\} = \mathrm{supp}_\rho(\{p, \delta_x\})$.*

Next, we impose Risk Aversion – noting that this is imposed here for purely technical reasons, as it is behaviorally distinct from deliberate stochasticity.[16] Consider two lotteries p and q such that q is a mean preserving spread of p, and suppose that the agent consistently picks q against some δ_x. Now suppose that we replace q with p, and δ_x with δ_y where $y < x$. We are making the unchosen option worse (as $y < x$); and if the agent is risk averse, since q is a mean preserving spread of p, we are also making the chosen option better. We thus posit that p should be chosen against δ_y.

Axiom 5 (Risk Aversion)**.** *For each $p, q \in \Delta$, if q is a mean preserving spread of p and $\{q\} = \mathrm{supp}_\rho(\{q, \delta_x\})$ for some $x \in (w,b]$, then $\{p\} = \mathrm{supp}_\rho(\{p, \delta_y\})$ for each $y \in [w,x)$.*

We are now ready to introduce the second representation in the paper. For this, denote the set of continuous functions from $[w,b]$ into $\mathbb{R}$ by $C([w,b])$ and metrize it by the supnorm. Given a lottery $p \in \Delta$ and a function $v \in C([w,b])$, we write $\mathbb{E}_p(v)$ for the Expected Utility of p with respect to v, that is, $\mathbb{E}_p(v) = \int_{[w,b]} v dp$.

Definition 3. A stochastic choice function ρ admits a Cautious Stochastic Choice representation if there exists a compact set $\mathcal{W} \subseteq C([w,b])$ such that every function $v \in \mathcal{W}$ is strictly increasing and concave and

$$\overline{\rho(A)} \in \arg\max_{p \in co(A)} \min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_p(v)), \forall A \in \mathcal{A}. \tag{1}$$

[16]This is needed because, as we already mentioned, we do not restrict how indifference is broken. Risk aversion guarantees the compactness of the set of utilities in our representation below, which guarantees strict monotonicity with respect to first order stochastic dominance, which, in turn, is essential to identify indifferences. Alternatively, we could replace risk aversion with a technical condition that guarantees compactness; we would obtain a similar representation but without the requirement that all functions are concave.

Theorem 3. *Let ρ be a stochastic choice function on Δ. The following statements are equivalent:*

(i) The stochastic choice function ρ satisfies Rational Mixing, Weak Stochastic Certainty Effect, Continuity, and Risk Aversion;

(ii) There exists a Cautious Stochastic Choice representation of ρ.

In a Cautious Stochastic Choice model, the agent has a *set* of utility functions $\mathcal{W}$, all of which are continuous, strictly increasing, and concave. It is as if she were unsure of which utility function to use to evaluate lotteries. She then proceeds as follows: for each lottery p, she computes the certainty equivalent with respect to every utility v in $\mathcal{W}$, and picks the *smallest* one. Note that if for some $u, v \in \mathcal{W}$ and $p, q \in \Delta$ we have $\mathbb{E}_p(u) > \mathbb{E}_q(u)$ but $\mathbb{E}_p(v) < \mathbb{E}_q(v)$, i.e., p is better for one utility but q is better for another, then the agent may prefer to mix p and q: this way, she obtains a lottery that is "not too bad" according to either u or v. This is similar to how, in the context of decision making under uncertainty, hedging may make an ambiguity averse agent better off. It is easy to see that these preferences are weakly convex, and – locally – may be strictly convex. The cautious criterion may thus generate a strict desire to mix existing options, leading to Deliberate Stochastic Choice.

The choice procedure the agent uses in the Cautious Stochastic Choice model is a special case of the Cautious Expected Utility model of Cerreia-Vioglio et al. (2015).[17] This is derived here by imposing Weak Stochastic Certainty Effect, which is reminiscent of NCI, which in turn characterizes the Cautious Expected Utility model. However, we have seen that Weak Stochastic Certainty Effect does not apply to the (unobserved) underlying preference, but constrains behavior only in extreme situations where the stochastic choice function is degenerate. It is thus conceptually much weaker than imposing NCI on the whole underlying preferences. The theorem above shows, however, that this is actually equivalent: within the context of stochastic choice, ruling out the opposite of Certainty Bias in such extreme cases is sufficient to guarantee the existence of a Cautious Expected Utility representation, and thus the whole underlying preferences abide by NCI. Put differently: within the context of stochastic choice, the Cautious Expected Utility model emerges by only restricting behavior on extreme cases.

The Cautious Stochastic Choice model is also conceptually related to models of Random Utility, where the agent has a probability distribution over possible utility functions and, for each decision, one utility is chosen randomly.[18] Here we also have multiple utilities, but it is as if the agent took into account *all* utilities at the same time – as if she were aware, or meta-cognitive, of this multiplicity – and reacted

[17]It is a special case in that $\mathcal{W}$ is compact and all utilities are concave. (The latter follows from risk aversion.)

[18]A formal comparison between these models appears in Section 5.

with caution, by using only the utility with the lowest certainty equivalent. That is, instead of using one random utility, only the most 'cautious' one is employed.[19]

4.1 Stochastic Choice and Certainty Bias

In the Cautious Stochastic Choice model, the underlying preferences of the agent follow Cautious Expected Utility, and it is the multiplicity of utilities together with the agent's caution that generates the desire to choose stochastically. On the other hand, Cautious Expected Utility was developed to address the Certainty Bias, where again the multiplicity of utilities and the caution generate this behavior: non-degenerate lotteries are evaluated with caution, while degenerate ones are not, since their certainty equivalent is the same no matter which utility is used. This suggests that our model of stochastic choice may entail a relation between the Certainty Bias and the stochasticity of choice. We will now formalize this intuition.

To do so, define when stochastic choice shows at least one instance of 'strict' Certainty Bias, where a strict advantage is given to certainty.

Definition 4. We say that a stochastic choice function ρ is *Certainty Biased* if there exist $A \in \mathcal{A}$, $x, y \in [w, b]$, with $x > y$, $r \in \Delta$, and $\lambda \in (0, 1)$ such that

$$\{\delta_y\} = \text{supp}_\rho(A \cup \{\delta_y\}) \quad \text{and} \quad \{\lambda\delta_x + (1-\lambda)r\} \neq \text{supp}_\rho(\lambda(A \cup \{\delta_x\}) + (1-\lambda)r).$$

Proposition 1. *Let ρ be a stochastic choice function that admits a Cautious Stochastic Choice representation. The following holds:*

1. *If ρ is a non-degenerate stochastic choice function, then ρ is Certainty Biased and all Cautious Stochastic Choice representations of ρ must have $|\mathcal{W}| > 1$.*
2. *If ρ is Certainty Biased and it admits a Cautious Stochastic Choice representation with $|\mathcal{W}| < \infty$, then ρ is a non-degenerate stochastic choice function and all Cautious Stochastic Choice representations $\mathcal{W}$ of ρ must have $|\mathcal{W}| > 1$.*

The proposition above shows a connection between Certainty Bias and stochasticity of choice. If ρ admits a representation with finitely many utilities, we observe stochastic choice in cases beyond those of indifference if, and only if, we observe at least one instance of strict Certainty Bias. These take place only if all representations involve more than one utility ($|\mathcal{W}| > 1$): it is the joint presence of multiple utilities as well as caution that leads to *both* stochastic choice as well as Certainty Bias.

When a representation of ρ contains infinitely many utilities, we may have Certainty Bias but ρ may not be non-degenerate. For example, if the preferences induced by $\mathcal{W}$ satisfy Betweenness, which implies linear indifference curves (and convex indifference sets), then ρ will never be non-degenerate, but it could well be Certainty

[19] The uniqueness properties and an elicitation procedure are discussed in Appendix B.

Biased (e.g., if the underlying preference follows Gul, 1991's model of disappointment aversion). However, as the proposition above shows, this is not possible if $|\mathcal{W}| < \infty$: it can be shown that in this case preferences must violate Betweenness, thus admitting areas of strict convexity, where non-degenerate stochastic choice can be found.

We note also that the link between the Certainty Bias and stochasticity suggested above finds experimental support in Agranov and Ortoleva (2017), where the stochasticity of answers is correlated with the tendency to exhibit Allais-like behavior.

5 Relation with Models in the Literature

Random Utility. As we discussed, Theorem 2 can be used to easily compare the Deliberate Stochastic Choice model with models of Random Utility. Formally, we say that a stochastic choice function ρ admits a Random Utility representation if there exists a probability measure over utilities such that for each alternative s in a choice problem A, the probability of choosing s from A, $\rho(A)(s)$, equals the probability of drawing a utility function u such that s maximizes u in A.[20]

It is well-known that a stochastic choice function that admits a Random Utility representation must satisfy Regularity. This is intuitive: if an option is the best according to one utility, its choice cannot be made less likely by removing alternatives. (In models of Random Utility, there is no complementarity between the chosen elements.) But then, following Theorem 2 we have a sharp distinction between our model and models of Random Utility: the only behavior that can be represented by both models is one of a *degenerate* Random Utility model – i.e., with only one utility possible – in which the agent exhibits stochastic behavior only when she is indifferent. Another immediate implication is that observing a violation of Regularity – an easily testable condition – implies that the agents' behavior cannot be represented by Random Utility, while it may be represented by Deliberate Stochastic Choice.

Random Expected Utility. Gul and Pesendorfer (2006) axiomatize the Random Expected Utility model, a version of Random Utility where all the utility functions involved are of the Expected Utility type. One of the conditions that characterize this model is Linearity:

Axiom 6 (Linearity)**.** *For each $A \in \mathcal{A}$, $p \in A$, $q \in \Delta$, and $\lambda \in (0,1)$,*

$$\rho(A)(p) = \rho(\lambda A + (1-\lambda)q)(\lambda p + (1-\lambda)q).$$

[20]Stochastic choice functions over a finite space of alternatives that admit a Random Utility representation were axiomatized by Falmagne (1978) (see also Barberá and Pattanaik, 1986). An issue arises when the utility functions allow for indifferences; assumptions are needed on how they are resolved. Two approaches have been suggested. First, to impose that the measure of the set of utility functions such that the maximum is not unique is zero for every choice problem. Second, to impose a tie-breaking rule that is independent of the choice problem.

We now show that if ρ admits a continuous Deliberate Stochastic Representation and in addition satisfies Linearity, then ρ is a *degenerate* Random Expected Utility model, i.e., again a model with only one utility. Formally:

Proposition 2. *Let ρ be a stochastic choice function that admits a continuous Deliberate Stochastic Choice representation and satisfies Linearity. Then, there exists a continuous function $u : [w, b] \to \mathbb{R}$ such that, for any choice problem A,*

$$\text{supp}_\rho(A) \subseteq \{p \in A : \mathbb{E}_p(u) \geq \mathbb{E}_q(u)\ \forall q \in A\}.$$

Deliberate Randomization. A small existing literature has suggested models of stochastic choice as deliberate randomization. As we have discussed, our model extends the intuition of Machina (1985) (see also Marley, 1997 and Swait and Marley, 2013) in a fully axiomatic setup.[21] Dwenger et al. (2016) propose a model in which agents choose to randomize following a desire to minimize regret. Their key assumption is that the regret after making the wrong choice is smaller if the choice is stochastic rather than deterministic.

Fudenberg et al. (2015) provide conditions under which stochastic choice corresponds to the maximization of Expected Utility and a perturbation function that depends only on the choice probabilities. Formally, they axiomatize a stochastic choice function ρ such that, for each choice problem A,

$$\rho(A) = \underset{p \in \Delta(A)}{\arg\max} \sum_{x \in A} [p(x)u(x) - c(p(x))], \tag{2}$$

where $\Delta(A)$ is the set of probability measures on A, u is a von Neumann-Morgenstern utility function and $c : [0, 1] \to \mathbb{R} \cup \{\infty\}$ is strictly convex and C^1 in $(0, 1)$. They call this representation *Weak Additive Perturbed Utility* (Weak APU).[22] Because of the strictly convex perturbation function c, this functional form gives the agent an intrinsic incentive to randomize. However, there are two important differences with our model.

A first difference is that we study a domain of menus of lotteries while Fudenberg et al. (2015) study menus of final outcomes. This is not a mere technical difference, as our goal is to study, in the spirit of Machina (1985), the link of stochastic choice with non-Expected Utility behavior – and preferences over lotteries must necessarily be present for a comparison to be possible.

---

[21] Machina (1985) suggests the following condition: if $A, A' \in \mathcal{A}$ are such that $co(A) \subset co(A')$ and $\overline{\rho(A')} \in co(A)$, then $\overline{\rho(A')} = \overline{\rho(A)}$. (This condition is related to Sen's α axiom.) While naturally related to our Rational Mixing axiom, this condition is not sufficient to characterize our model. (Unless preferences are strictly convex, it is also not necessary, because of indifferences: for example, A and A' may differ only for the inclusion of some strictly dominated option that is never chosen in either case, but the stochastic choice may not coincide as indifference may be resolved differently.)

[22] Their paper also characterizes the case in which the function c satisfy the additional requirement that $\lim_{q \to 0} c'(q) = -\infty$, which they call an Additive Perturbed Utility representation.

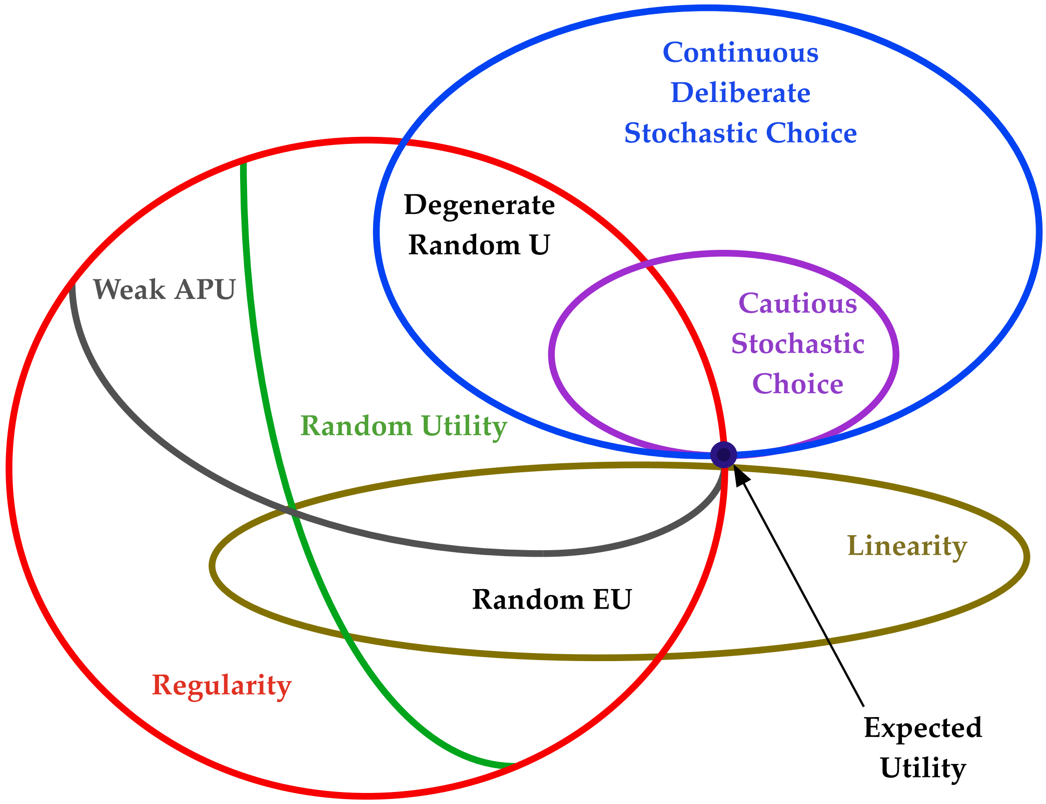

Figure 1: Relation with other models

A second, crucial difference between the models is that even though the model in Fudenberg et al. (2015) rewards probabilistic choices and this sometimes gives the individual an incentive to randomize, their model *does* satisfy Regularity (Fudenberg et al., 2015, p. 2386). This is a crucial conceptual difference, as it implies that their model does not include one of the main driving forces of ours, as we discussed above. It also implies that the formal relation between their models and ours is the same as with Random Utility: the only behavior compatible with both is one of an agent that exhibits stochastic choice only when indifferent.[23]

The results above are summarized in Figure 1.

Other Related Literature. This paper is related to various strands besides stochastic choice. First, our work is also related to the literature on non-Expected Utility, since a desire to randomize would emerge only as long as the underlying preferences

[23]An alternative way to apply their paper to the case of lotteries is, instead of using their representation theorem directly, to use their functional form, $\sum[p(x)u(x) - c(p(x))]$, as a representation for the preferences in our Theorem 1. This would lead to a model that is a hybrid of the two formulations.

over lotteries are, at least on some region, strictly convex in the probabilities, in violation of Expected Utility. This cannot be the case, for example, in the Rank Dependent Expected Utility model of Quiggin (1982) if distortions are pessimistic: in this case subjects are *averse* to mixing. (The converse would hold if they were optimistic, or in some areas when distortions are inverted S-shaped.) No such preference would emerge also in the case of the Disappointment Aversion model of Gul (1991), as well as in any other member of the Betweenness class, which implies indifference to randomization. The class of convex preferences over lotteries was studied in Cerreia-Vioglio (2009). Our Cautious Stochastic Choice model includes as a representation of the underlying preferences a special case of the Cautious Expected Utility model of Cerreia-Vioglio et al. (2015). As previously mentioned, our analysis here differs as we cannot observe the preferences, and thus impose a postulate reminiscent of NCI but only for the extreme case in which the stochastic choice function is degenerate.

Stoye (2015) studies choice environments in which agents can randomize at will (thus restricting observability to convex sets). Considering as a primitive the choice correspondence of the agent in an Anscombe-Aumann setup,[24] he characterizes various models of choice under uncertainty that include a desire to randomize. Unlike Stoye, we take as a primitive the agent's stochastic choice function, instead of the choice correspondence; this not only suggests different interpretations, but also entails substantial technical differences. In addition, we study a setup with risk, and not uncertainty, and characterize the most general model of deliberate randomization given a complete preference relation over monetary lotteries.

As we have mentioned, our general representation theorem (Theorem 1) is related to the literature on revealed preferences on finite datasets. By randomizing over a set of alternatives, the agent can obtain any point of its convex hull. It is as if we could only see individuals' choices from convex sets, restricting our ability to observe the entire preferences. Our problem is then related to the issue of eliciting preferences with limited datasets, originated by Afriat (1967), and for our first theorem we employ techniques from this literature. Our results are particularly related to Chambers and Echenique (2016) and Nishimura et al. (2017).

[24]The paper considers also a setup with pure risk, but in that case the analysis is mostly focused on characterizing the case of Expected Utility, where there is no desire to randomize.

Appendix A: Preliminary Results

In this section we present a result that extends the analysis of Cerreia-Vioglio et al. (2015) which will be instrumental to prove the results in the second part of this paper. For that, as we did in the main text, let $[w, b]$ be a closed non-trivial interval in $\mathbb{R}$ and let Δ be the space of Borel probability measures on $[w, b]$ endowed with the topology of weak convergence. Our primitive will be a binary relation $\succsim$ on Δ. We will impose the following postulates on $\succsim$:

Axiom 7 (Weak Order)**.** *The relation $\succsim$ is complete and transitive.*

Axiom 8 (Continuity)**.** *For each $q \in \Delta$, the sets $\{p \in \Delta : p \succsim q\}$ and $\{p \in \Delta : q \succsim p\}$ are closed.*

Axiom 9 (Monotonicity)**.** *For each $x, y \in [w, b]$ and $\lambda \in (0, 1]$,*

$$x > y \;\Rightarrow\; \lambda\delta_x + (1-\lambda)\delta_w \succ \lambda\delta_y + (1-\lambda)\delta_w.$$

Axiom 10 (Negative Certainty Independence)**.** *For each $p, q \in \Delta$, $x \in [w, b]$ and $\lambda \in [0, 1]$,*

$$p \succsim \delta_x \;\Rightarrow\; \lambda p + (1-\lambda)q \succsim \lambda\delta_x + (1-\lambda)q.$$

Axiom 11 (Risk Aversion)**.** *For each $p, q \in \Delta$, if q is a mean preserving spread of p, then $p \succsim q$.*

We can now state the following theorem.

Theorem 4. *Let $\succsim$ be a binary relation on Δ. The following statements are equivalent:*

(i) The relation $\succsim$ satisfies Weak Order, Continuity, Monotonicity, Negative Certainty Independence, and Risk Aversion.

(ii) There exists a compact set $\mathcal{W} \subseteq C([w, b])$ such that every function $v \in \mathcal{W}$ is strictly increasing and concave and, for every $p, q \in \Delta$,

$$p \succsim q \iff \min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_p(v)) \geq \min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_q(v)).$$

Let $\mathcal{U}$ be the set of strictly increasing and continuous functions from $[w, b]$ into $\mathbb{R}$. Define $\mathcal{U}_{nor}$ by $\mathcal{U}_{nor} = \{v \in \mathcal{U} : v(w) = 0 \text{ and } v(b) = 1\}$. We first prove two auxiliary results (Lemma 1 and Theorem 5 below).

Lemma 1. *Let $\mathcal{W}$ be a subset of $\mathcal{U}_{nor}$. The following statements are equivalent:*

(i) $\mathcal{W}$ is compact with respect to the topology of sequential pointwise convergence;

(ii) $\mathcal{W}$ is norm compact.

Proof of Lemma 1. It is trivial that (ii) implies (i). For the other direction, consider $(v_n) \in \mathcal{W}^\infty$. Observe that, by construction, (v_n) is uniformly bounded. By assumption, there exists a subsequence (v_{n_k}) of (v_n) and $v \in \mathcal{W}$ such that $v_{n_k}(x) \to v(x)$ for all $x \in [w, b]$. By Aliprantis and Burkinshaw (1998, p. 79) and since v is a continuous function and each v_{n_k} is increasing, it follows that this convergence is uniform, proving the statement. ■

Theorem 5. *Let $V : \Delta \to \mathbb{R}$ and $\mathcal{W} \subseteq \mathcal{U}_{nor}$ be such that*

$$V(p) = \inf_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_p(v)) \qquad \forall p \in \Delta.$$

If each element of $\mathcal{W}$ is concave, V is continuous, and such that for each $x, y \in [w, b]$ and for each $\lambda \in (0, 1]$

$$x > y \Longrightarrow V(\lambda\delta_x + (1-\lambda)\delta_w) > V(\lambda\delta_y + (1-\lambda)\delta_w) \tag{3}$$

then $\mathcal{W}$ is relatively compact with respect to the topology of sequential pointwise convergence restricted to $\mathcal{U}_{nor}$.

Proof of Theorem 5. Let us first show that, for each $\varepsilon > 0$, there exists $\delta \in (0, b-w)$ such that, for each $v \in \mathcal{W}$,

$$v(w+\delta) < \varepsilon. \tag{4}$$

By contradiction, assume that there exists $\bar{\varepsilon} > 0$ such that for each $\delta \in (0, b-w)$ there exists $v_\delta \in \mathcal{W}$ such that $v_\delta(w+\delta) \geq \bar{\varepsilon}$. In particular, for each $k \in \mathbb{N}$ such that $\frac{1}{k} < b - w$ there exists $v_k \in \mathcal{W}$ such that $v_k\left(w + \frac{1}{k}\right) \geq \bar{\varepsilon}$. Define $\lambda_k \in [0, 1]$ for each $k > \frac{1}{b-w}$ to be such that

$$\lambda_k v_k(b) + (1-\lambda_k) v_k(w) = \lambda_k = v_k\left(w + \frac{1}{k}\right) \geq \bar{\varepsilon} > 0. \tag{5}$$

Define $p_k = \lambda_k\delta_b + (1-\lambda_k)\delta_w$ for all $k > \frac{1}{b-w}$. Without loss of generality, we can assume that $\lambda_k \to \lambda$. Notice that $\lambda \geq \bar{\varepsilon}$. Define $p = \lambda\delta_b + (1-\lambda)\delta_w$. It is immediate to see that $p_k \to p$. By (5) and by definition of V, it follows that

$$w \leq V(p_k) \leq v_k^{-1}(\mathbb{E}_{p_k}(v_k)) = w + \frac{1}{k} \qquad \forall k > \frac{1}{b-w}.$$

Since V is continuous and by passing to the limit, we have that

$$V(\lambda\delta_b + (1-\lambda)\delta_w) = V(p) = w = V(\lambda\delta_w + (1-\lambda)\delta_w),$$

a contradiction with V satisfying (3). Now, consider $(v_n) \in \mathcal{W}^\infty$. Observe that, by construction, (v_n) is uniformly bounded. By Rockafellar (1970, Theorem 10.9), there exists a subsequence (v_{n_k}) of (v_n) and $v \in \mathbb{R}^{(w,b)}$ such that $v_{n_k}(x) \to v(x)$ for all $x \in (w,b)$. Since $v_{n_k}([w,b]) = [0,1]$ for all $k \in \mathbb{N}$, v takes values in $[0,1]$. Define $\bar{v} : [w,b] \to [0,1]$ by

$$\bar{v}(w) = 0,\ \bar{v}(b) = 1, \text{ and } \bar{v}(x) = v(x) \qquad \forall x \in (w,b).$$

Since $v_{n_k} \in \mathcal{W} \subseteq \mathcal{U}_{nor}$ for every $k \in \mathbb{N}$, we have that $v_{n_k}(x) \to \bar{v}(x)$ for all $x \in [w,b]$. It is immediate to see that $\bar{v}$ is increasing and concave. We are left to show that $\bar{v} \in \mathcal{U}_{nor}$, that is, $\bar{v}$ is continuous and strictly increasing. By Rockafellar (1970, Theorem 10.1) and since $\bar{v}$ is finite and concave, we have that $\bar{v}$ is continuous at each point of (w,b). We are left to check continuity at the extrema. Since $\bar{v}$ is increasing, concave, and such that $\bar{v}(w) = 0 = \bar{v}(b) - 1$, we have that $\bar{v}(x) \geq \frac{x-w}{b-w}$ for all $x \in [w,b]$. It follows that $1 \geq \limsup_{x \to b^-} \bar{v}(x) \geq \liminf_{x \to b^-} \bar{v}(x) \geq \lim_{x \to b^-} \frac{x-w}{b-w} = 1$, proving continuity at b. We next show that $\bar{v}$ is continuous at w. By (4), for each $\varepsilon > 0$ we have that there exists $\delta > 0$ such $v_{n_k}(w+\delta) < \frac{\varepsilon}{2}$ for all $k \in \mathbb{N}$. Since $\bar{v}(w) = 0$, $\bar{v}$ is increasing, and the pointwise limit of (v_{n_k}), we have that for each $x \in [w, w+\delta)$

$$\begin{aligned} |\bar{v}(x) - \bar{v}(0)| &= |\bar{v}(x)| \leq \bar{v}(x) \leq \bar{v}(w+\delta) \\ &= \lim_k v_{n_k}(w+\delta) \leq \frac{\varepsilon}{2} < \varepsilon, \end{aligned}$$

proving continuity at w. We are left to show that $\bar{v}$ is strictly increasing. We argue by contradiction. Assume that $\bar{v}$ is not strictly increasing. Since $\bar{v}$ is increasing, continuous, concave, and such that $\bar{v}(w) = 0 = \bar{v}(b) - 1$, there exists $x \in (w,b)$ such that $\bar{v}(x) = 1$. Define $(\lambda_k) \in [0,1)^\infty$ to be such that $\lambda_k v_{n_k}(b) + (1 - \lambda_{n_k}) v_{n_k}(w) = \lambda_k = v_{n_k}(x)$. Since $\bar{v}$ is the pointwise limit of (v_{n_k}), it follows that $\lambda_k \to 1$. Define $p_k = \lambda_k \delta_b + (1-\lambda_k)\delta_w$ for all $k \in \mathbb{N}$. It is immediate to see that $p_k \to \delta_b$. Thus, we also have that

$$V(p_k) \leq v_{n_k}^{-1}(\mathbb{E}_{p_k}(v_{n_k})) \leq x.$$

Since V is continuous and by passing to the limit, we have that $x < b = V(\delta_b) \leq x$, a contradiction. ■

Proof of Theorem 4. (i) implies (ii). By Cerreia-Vioglio et al. (2015, Theorem 2) and since Monotonicity implies Weak Monotonicity, there exists $\mathcal{W} \subseteq \mathcal{U}_{nor}$ such that the function $V : \Delta \to \mathbb{R}$, defined by

$$V(p) = \inf_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_p(v)) \qquad \forall p \in \Delta, \tag{6}$$

is a continuous utility function for $\succsim$. By Cerreia-Vioglio et al. (2015, Theorem 3), each $v \in \mathcal{W}$ is concave and $\mathcal{W}$ can be chosen to be closed under the topology of sequential pointwise convergence restricted to $\mathcal{U}_{nor}$ (it is enough to consider the convex set $\mathcal{W}_{\max-\text{nor}}$ in Cerreia-Vioglio et al. 2015, Proof of Step 2, Theorem 1). Since $\succsim$ satisfies Monotonicity, V is such that for each $x, y \in [w,b]$ and for each $\lambda \in (0,1]$

$$x > y \Rightarrow V(\lambda\delta_x + (1-\lambda)\delta_w) > V(\lambda\delta_y + (1-\lambda)\delta_w).$$

By Theorem 5, it follows that $\mathcal{W}$ is in fact compact under the topology of sequential pointwise convergence restricted to $\mathcal{U}_{nor}$. By Lemma 1, this implies that $\mathcal{W}$ is also compact with respect to the topology induced by the supnorm. We can conclude that the inf in (6) is attained and so the statement follows.

(ii) implies (i). Consider $V : \Delta \to \mathbb{R}$ defined by

$$V(p) = \min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_p(v)) \qquad \forall p \in \Delta.$$

By hypothesis, V is well defined and it represents $\succsim$. Since $\mathcal{W}$ is compact, we have that V is continuous. By Cerreia-Vioglio et al. (2015, Theorems 1–3), $\succsim$ satisfies Weak Order, Continuity, Negative Certainty Independence and Risk Aversion. Next, consider $p, q \in \Delta$ such that $p \succ_{FSD} q$. Consider also $v \in \mathcal{W}$ such that $V(p) = v^{-1}(\mathbb{E}_p(v))$. Since v is strictly increasing, we have that $V(p) = v^{-1}(\mathbb{E}_p(v)) > v^{-1}(\mathbb{E}_q(v)) \geq V(q)$, proving that $\succsim$ strictly preserves first order stochastic dominance and so, in particular, Monotonicity. ■

Appendix B: Additional Results

Continuous Deliberate Stochastic Choice Models

In this section we extend the result of Theorem 1 to the case in which the underlying preference relation admits a representation by a continuous utility function. For this we need to strengthen the consistency condition in the Rational Mixing axiom to apply not only to the transitive closure of the relation R defined in Remark 2, but to its closed transitive closure. Formally, define the binary relation R on Δ as

$$pRq \text{ iff } \exists A \in \mathcal{A} \text{ s.t. } p = \overline{\rho(A)} \text{ and } q \in co(A).$$

Intuitively, pRq if it ever happens that p is chosen, either directly ($\{p\} = \text{supp}_\rho(A)$) or as the outcome of randomization ($p = \overline{\rho(A)}$), from a set A where q could have also been chosen ($q \in co(A)$). Denote by $\overline{tran(R)}$ the minimal continuous and transitive binary relation on Δ such that $R \subseteq \overline{tran(R)}$. We call $\overline{tran(R)}$ the closed transitive closure of R. We can now write the following postulate:

Axiom 12 (Continuous Rational Mixing)**.** *If $p, q \in \Delta$ are such that $p\overline{tran(R)}q$, then it cannot be the case that $q >_{FOSD} p$.*

We have the following result:

Proposition 3. *A stochastic choice function ρ satisfies Continuous Rational Mixing if and only if it admits a Deliberate Stochastic Choice representation $\succsim$ that can be represented by a continuous utility function.*

Proof of Proposition 3. Suppose first that ρ admits a Deliberate Stochastic Choice representation $\succsim$ that can be represented by a continuous utility function. This implies that $\succsim$ is a continuous preorder. Since, by the representation of ρ, $R \subseteq \succsim$, this implies that $\overline{tran(R)} \subseteq \succsim$. But then, if $p, q \in \Delta$ are such that $p\overline{tran(R)}q$, we also have that $p \succsim q$, which implies that it is not true that $q >_{FOSD} p$. That is, ρ satisfies Continuous Rational Mixing.

Conversely, suppose ρ is a stochastic choice function that satisfies Continuous Rational Mixing. Pick any pair of lotteries p and q in Δ with $q >_{FOSD} p$. This implies that $q >_{FOSD} \alpha q + (1-\alpha)p$ for every $\alpha \in [0,1)$. By Continuous Rational Mixing, we must have that $\rho(\{p,q\}) = q$, which implies that qRp. Moreover, again by Continuous Rational Mixing, we cannot have that $p\overline{tran(R)}q$. This shows that $\overline{tran(R)}$ is an extension of the first order stochastic dominance relation. By Levin's Theorem, there exists a continuous function $u : \Delta \to \mathbb{R}$ such that $p\overline{tran(R)}q$ implies $u(p) \geq u(q)$, with strict inequality whenever it is not true that $q\overline{tran(R)}p$. Now we can proceed as in the proof of Theorem 1, using the preference relation the function u induces, to conclude the proof. ■

Uniqueness properties of the Cautious Stochastic Choice model

We now discuss the uniqueness properties of the Cautious Stochastic Choice model. First of all, the relation $\succsim$ induced by a Cautious Stochastic Choice representation is unique.

Proposition 4. *Consider two Cautious Stochastic Choice representations $\mathcal{W}$ and $\mathcal{W}'$ of some stochastic choice function ρ. Then, for each $q \in \Delta$*

$$\min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_q(v)) = \min_{v \in \mathcal{W}'} v^{-1}(\mathbb{E}_q(v)).$$

Proof of Proposition 4. In order to prove the proposition, we will show that for every Cautious Stochastic Choice representation $\mathcal{W}$ of a stochastic choice function ρ we have that, for every $q \in \Delta \backslash \{\delta_b, \delta_w\}$,

$$\min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_q(v)) = \sup\{y \in [w,b] : \{q\} = \operatorname{supp}_\rho(\{q, \delta_y\})\}.$$

To see that, first notice that it is clear from the fact that $\mathcal{W}$ is a Cautious Stochastic Choice representation of ρ that $\{q\} = \operatorname{supp}_\rho(\{q, \delta_y\})$ for every $y \in [w,b]$ such that

$$\min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_q(v)) > y,$$

and that $\{\delta_y\} = \operatorname{supp}_\rho(\{q, \delta_y\})$ for every $y \in [w,b]$ such that

$$\min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_q(v)) < y.$$

This can happen only if

$$\min_{v\in\mathcal{W}} v^{-1}(\mathbb{E}_q(v)) = \sup\{y \in [w,b] : \{q\} = \text{supp}_\rho(\{q,\delta_y\})\},$$

proving the statement. ■

Less straightforward is the uniqueness of $\mathcal{W}$. These are similar uniqueness properties as in Cerreia-Vioglio et al. (2015), to which we refer for further detail. Suppose that $\mathcal{W}$ is a Cautious Stochastic Choice representation of a given stochastic choice function ρ. First, we can normalize all functions $v \in \mathcal{W}$ so that $v(w) = 0$ and $v(b) = 1$; call this a *normalized* Cautious Stochastic Choice model. Second, the convex hull of $\mathcal{W}$, $co(\mathcal{W})$, would represent the same preferences. Lastly, we can always add redundant functions to the set without changing the representation, like a $\bar{v}$ that is a continuous, strictly increasing, and strictly convex transformation of some $v \in \mathcal{W}$. To obtain uniqueness, we have to remove these redundant functions and aim to obtain a "minimal" set. Our next result establishes that there exists a Cautious Stochastic Choice representation with a minimal, normalized, and convex set of utilities.

Proposition 5. *Let ρ be a stochastic choice function that admits a Cautious Stochastic Choice representation. Then, there exists a normalized and convex Cautious Stochastic Choice representation $\widehat{\mathcal{W}}$ of ρ such that, for any other normalized and convex Stochastic Choice representation of ρ, we have $\widehat{\mathcal{W}} \subseteq \mathcal{W}$.*

Proof of Proposition 5. By Proposition 4, any two Cautious Stochastic Choice representations, $\mathcal{W}$ and $\mathcal{W}'$, of the same stochastic choice function ρ represent the same binary relation $\succsim$ via the utility

$$V(p) = \min_{v\in\mathcal{W}} v^{-1}(\mathbb{E}_p(v)) \qquad \forall p \in \Delta. \tag{7}$$

By the proof of Theorem 4, we also know that V in (7) is also represented by the normalized, compact and convex set $\mathcal{W}_{\max-\text{nor}}$. By Theorem 2 (and its proof) in Cerreia-Vioglio et al. (2015) and since $\mathcal{W}$ is convex and compact, $\mathcal{W}_{\max-\text{nor}} \subseteq \text{cl}(\text{co}(\mathcal{W})) = \text{cl}(\mathcal{W}) = \mathcal{W}$. By setting $\widehat{\mathcal{W}} = \mathcal{W}_{\max-\text{nor}}$ we obtain the desired conclusion. ■

We conclude this discussion by suggesting a method that can be used by experimental and applied researchers to identify from choice data the set of utilities of a Cautious Stochastic Choice model. Assume that our data can be described by a Cautious Stochastic Choice representation $\mathcal{W}$, and consider an arbitrary concave and strictly increasing function $v^* \in C([w,b])$. Now suppose that $x \in [w,b]$ and $p \in \Delta$ are such that $p \in \text{supp}_\rho(\{p,\delta_x\})$. By the representation of ρ, this can happen only if

$$\min_{v\in\mathcal{W}} v^{-1}(\mathbb{E}_p(v)) \geq x.$$

But then, if $\mathbb{E}_p(v^*) < v^*(x)$, we can be sure that v^* does not belong to $\mathcal{W}$. This discussion can be summarized by the following proposition:

Proposition 6. *Let ρ be a stochastic choice function that admits a Cautious Stochastic Choice representation and let $v^* \in C([w,b])$ be strictly increasing and concave. If there exist $x \in [w,b]$ and $p \in \Delta$ such that $p \in \text{supp}_\rho(\{p, \delta_x\})$ and $\mathbb{E}_p(v^*) < v^*(x)$, then $v^* \notin \mathcal{W}$ for every Cautious Stochastic Choice representation $\mathcal{W}$ of ρ.*

This result shows how utilities can be easily excluded from the set $\mathcal{W}$, helping the identification of the set. Specifically, to exclude a function v^*, all one needs to do is to: 1) find a lottery p and a certain amount x such that $p \in \text{supp}_\rho(\{p, \delta_x\})$; 2) compare the magnitudes of $\mathbb{E}_p(v^*)$ and $v^*(x)$. The proposition shows that if $v^*(x)$ is strictly greater than $\mathbb{E}_p(v^*)$, then v^* cannot belong to $\mathcal{W}$.

Appendix C: Proof of the Results in the Main Text

Proof of Theorem 1. It is clear that if ρ admits a Deliberate Stochastic Choice representation, then ρ satisfies Rational Mixing. Suppose, thus, that ρ satisfies Rational Mixing and define the binary relation R the same way it is defined in the main text (see Remark 2). Pick any pair of lotteries p and q such that $p >_{FOSD} q$. This implies that $p >_{FOSD} (\alpha p + (1-\alpha) q)$ for every $\alpha \in [0,1)$. Define $A_1 = \{p,q\}$ and $A_2 = \{p\}$. Notice that Rational Mixing implies that we must have $\overline{\rho(A_1)} = p$. Consequently, we have pRq. Moreover, if we have $k \in \mathbb{N}$ and $A_1, ..., A_k$ such that $\overline{\rho(A_1)} = q$ and $\overline{\rho(A_i)} \in co(A_{i-1})$ for $i = 2, ..., k$, Rational Mixing implies that $p \notin co(A_k)$. This shows that we cannot have $q\,tran(R)p$. We conclude that $tran(R)$ is an extension of the first order stochastic dominance relation. Now pick any complete extension $\succsim$ of $tran(R)$. By what we have just seen, $\succsim$ is also an extension of the first order stochastic dominance relation. Moreover, by definition, we have that $\overline{\rho(A)}Rq$ for every $q \in co(A)$, for every $A \in \mathcal{A}$. Consequently, we have $\overline{\rho(A)} \succsim q$ for every $q \in co(A)$, for every $A \in \mathcal{A}$. This proves Theorem 1. ■

Proof of Theorem 2. (i) implies (iii). Assume that ρ is non-degenerate. Then, there exist $p, q \in \Delta$ with $|\text{supp}_\rho(\{p,q\})| \neq 1$ and $\lambda \in (0,1)$ such that

$$|\text{supp}_\rho(\{\lambda p + (1-\lambda)\delta_w, q\})| \neq 1 \quad \text{and} \quad |\text{supp}_\rho(\{p, \lambda q + (1-\lambda)\delta_w\})| \neq 1.$$

Since $\succsim$ is a Deliberate Stochastic Choice representation of ρ, we can assume without loss of generality that $p \succsim q$. Since $|\text{supp}_\rho(\{p,q\})| \neq 1$ and $\succsim$ strictly preserves first order stochastic dominance, note also that $p \neq q$ and so $p \succ \delta_w$. Since $|\text{supp}_\rho(\{p, \lambda q + (1-\lambda)\delta_w\})| \neq 1$, note also that there exists $\gamma \in (0,1)$ such that $\gamma(\lambda q + (1-\lambda)\delta_w) + (1-\gamma)p \succsim p$. It follows that

$$\begin{aligned} & \gamma(\lambda q + (1-\lambda)\delta_w) + (1-\gamma)p \\ = \; & [1-\gamma(1-\lambda)]\left(\frac{\gamma\lambda}{1-\gamma(1-\lambda)}q + \frac{1-\gamma}{1-\gamma(1-\lambda)}p\right) + \gamma(1-\lambda)\delta_w \\ = \; & (1-\beta)(\mu q + (1-\mu)p) + \beta\delta_w \end{aligned}$$

where $\mu = \frac{\gamma\lambda}{1-\gamma(1-\lambda)} \in (0,1)$ and $\beta = \gamma(1-\lambda) \in (0,1)$. Since $\succsim$ strictly preserves first order stochastic dominance and $p \succ \delta_w$, we have that $\mu q + (1-\mu)p \succ \delta_w$ and

$$\mu q + (1-\mu)p \succ (1-\beta)(\mu q + (1-\mu)p) + \beta\delta_w \succsim p \succsim q.$$

We can conclude that $\mu q + (1-\mu)p \succ p, q$.

(iii) implies (i). By assumption, there exist $p, q \in \Delta$ and $\lambda \in (0,1)$ such that

$$\lambda p + (1-\lambda)q \succ p, q. \tag{8}$$

Define $A = \{p, q\}$. It follows that there exist $\mu \in [0,1]$ such that $\mu p + (1-\mu)q = \overline{\rho(A)} \succsim r$ for all $r \in co(A)$. By (8), it follows that $\mu \in (0,1)$ and $\mu p + (1-\mu)q \succ p, q$, yielding that $|\text{supp}_\rho(\{p,q\})| \neq 1$. Without loss of generality, we can assume that $p \succsim q$. Since $\succsim$ is continuous, it follows that there exists $\gamma \in (0,1)$ such that

$$\gamma(\mu p + (1-\mu)q) + (1-\gamma)\delta_w \succ p \succsim q.$$

Note that

$$\begin{aligned}[1-\gamma(1-\mu)]\left(\frac{\gamma\mu}{1-\gamma(1-\mu)}p + \frac{1-\gamma}{1-\gamma(1-\mu)}\delta_w\right) + \gamma(1-\mu)q \\ = \gamma(\mu p + (1-\mu)q) + (1-\gamma)\delta_w.\end{aligned}$$

So,

$$(1-\beta)(\alpha p + (1-\alpha)\delta_w) + \beta q \succ p \succsim q$$

where $\beta = \gamma(1-\mu) \in (0,1)$ and $\alpha = \frac{\gamma\mu}{1-\gamma(1-\mu)} \in (0,1)$. Similarly, we have that

$$[1-\gamma\mu]\left(\frac{\gamma(1-\mu)}{1-\gamma\mu}q + \frac{1-\gamma}{1-\gamma\mu}\delta_w\right) + \gamma\mu p = \gamma(\mu p + (1-\mu)q) + (1-\gamma)\delta_w,$$

that is,

$$(1-\beta')(\alpha' q + (1-\alpha')\delta_w) + \beta' p \succ p \succsim q$$

where $\beta' = \gamma\mu \in (0,1)$ and $\alpha' = \frac{\gamma(1-\mu)}{1-\gamma\mu} \in (0,1)$. Next, define $\hat{\alpha} = \max\{\alpha, \alpha'\} \in (0,1)$. Since $\succsim$ strictly preserves first order stochastic dominance, we have that

$$(1-\beta)(\hat{\alpha} p + (1-\hat{\alpha})\delta_w) + \beta q \succ p \succsim q$$

and

$$(1-\beta')(\hat{\alpha} q + (1-\hat{\alpha})\delta_w) + \beta' p \succ p \succsim q.$$

We can conclude that

$$|\text{supp}_\rho(\{\hat{\alpha}p + (1-\hat{\alpha})\delta_w, q\})| \neq 1 \quad \text{and} \quad |\text{supp}_\rho(\{p, \hat{\alpha}q + (1-\hat{\alpha})\delta_w\})| \neq 1.$$

Since $\hat{\alpha} \in (0,1)$, ρ is non-degenerate.

(iii) implies (ii). Assume that there exist $p, q \in \Delta$ and $\lambda \in (0,1)$ such that $\lambda p + (1-\lambda)q \succ p, q$. Since $\succsim$ strictly preserves first order stochastic dominance, we must have $p \neq \delta_w$ and $q \neq \delta_w$. By continuity of $\succsim$, there exist maximal and minimal α^M and α_m in $[0,1]$ such that $\alpha^M p + (1-\alpha^M)q \sim \alpha_m p + (1-\alpha_m)q \succsim \alpha p + (1-\alpha)q$ for every $\alpha \in [0,1]$. Note that we must have $0 < \alpha_m \leq \alpha^M < 1$. By construction, $\alpha^M p + (1-\alpha^M)q \succ \lambda p + (1-\lambda)(\alpha^M p + (1-\alpha^M)q)$ for every $\lambda \in [\alpha_m, 1]$. Continuity of $\succsim$ and the fact that $\succsim$ strictly preserves first order stochastic dominance now imply that there exists $\varepsilon \in (0,1)$ such that

$$\begin{aligned} \varepsilon\delta_w + (1-\varepsilon)(\alpha^M p + (1-\alpha^M)q) &\succ \lambda p + (1-\lambda)(\alpha^M p + (1-\alpha^M)q) \\ &\succsim \lambda p + (1-\lambda)[\varepsilon\delta_w + (1-\varepsilon)(\alpha^M p + (1-\alpha^M)q)] \end{aligned}$$

for all $\lambda \in [\alpha_m, 1]$. Let $r = \varepsilon\delta_w + (1-\varepsilon)(\alpha^M p + (1-\alpha^M)q)$ and fix any $\hat{\rho}$ consistent with $\succsim$. The observation above implies that $\hat{\rho}(\{p, r\})(p) < \alpha_m$. However, the definition of α_m and the fact that $\succsim$ strictly preserves first order stochastic dominance imply that $\text{supp}_{\hat{\rho}}(\{p, q, r\}) \subseteq \{p, q\}$ and $\hat{\rho}(\{p, q, r\})(p) \geq \alpha_m$, which is a violation of Regularity.

(ii) implies (iii). By contradiction, assume that $\succsim$ does not have a point of strict convexity. That is, suppose that for all $p, q \in \Delta$ with $p \succsim q$, we have $p \succsim \lambda p + (1-\lambda)q$ for every $\lambda \in [0,1]$. Since $\succsim$ is a complete preorder, this is equivalent to say that $\succsim$ has convex lower contour sets. That is, the set

$$\{p \in \Delta : r \succsim p\}$$

is convex for all $r \in \Delta$. Now, let $\geq$ be any linear order (a complete, transitive and antisymmetric binary relation) on Δ (the existence of $\geq$ is guaranteed by the Well Ordering Principle, for example). Define $\trianglerighteq$ to be the relation that applies $\succsim$ and $\geq$ lexicographically. That is, for every $p, q \in \Delta$, $p \trianglerighteq q$ if and only if $p \succ q$ or $p \sim q$ and $p \geq q$. Note that $\trianglerighteq$ is also a linear order on Δ. Finally, let $\hat{\rho}$ be the stochastic choice function that, for each $A \in \mathcal{A}$, assigns probability one to the unique maximizer of $\trianglerighteq$ in A. It is clear that $\hat{\rho}$ satisfies Regularity. Moreover, by the definition of $\trianglerighteq$, if $p \in A$ is such that $\{p\} = \text{supp}_{\hat{\rho}}(A)$, then $p \succsim q$ for every $q \in A$. Since $\succsim$ has convex lower contour sets, this implies that, in fact, $p \succsim q$ for every $q \in co(A)$. That is, $\hat{\rho}$ is a stochastic choice function consistent with $\succsim$ that satisfies Regularity, which is a contradiction.

Proof of Theorem 3. Suppose first that ρ satisfies all the axioms in the statement of the theorem. By Theorem 1, ρ admits a Deliberate Stochastic Choice representation $\succsim$. We first need the following claim:

Claim 1. *For every $p \in \Delta$ and $x, y \in [w, b]$ with $x > y$, if $\{p\} = \text{supp}_\rho(\{p, \delta_x\})$, then $\{p\} = \text{supp}_\rho(\{p, \delta_y\})$.*

Proof of Claim. Since $\{p\} = \text{supp}_\rho(\{p, \delta_x\})$, we must have that $p \succsim \lambda p + (1-\lambda)\delta_x$ for every $\lambda \in [0,1]$. Since $\succsim$ extends the first order stochastic dominance relation, this implies that $p \succ \lambda p + (1-\lambda)\delta_y$ for every $\lambda \in [0,1)$. This now implies that $\{p\} = \text{supp}_\rho(\{p, \delta_y\})$. ‖

Now, for every $p \in \Delta$, define the set D_p by

$$D_p = \{x \in [w,b] : \{p\} = \text{supp}_\rho(\{p, \delta_x\})\}.$$

Claim 1 implies that D_p is an interval, for every $p \in \Delta$. That is, for every $p \in \Delta$, if $x \in D_p$, then $[w,x] \subseteq D_p$. Now define the function $V : \Delta \to [w,b]$ by $V(p) = \sup D_p$ for every $p \in \Delta$.[25] Notice that, since ρ admits a Deliberate Stochastic Choice representation, we must have $V(\delta_x) = x$ for every $x \in [w,b]$. We now need the following claim:

Claim 2. *For every choice problem A, $V(\overline{\rho(A)}) \geq V(q)$ for every $q \in co(A)$.*

Proof of Claim. Fix a choice problem A and $q \in co(A)$. Let $p = \overline{\rho(A)}$. If $V(q) = w$, then there is nothing to prove, so suppose that $V(q) > w$ and pick any $x, y, z \in (w, V(q))$ with $x > y > z$. By the definition of V and Claim 1, we have $\{q\} = \text{supp}_\rho(\{q, \delta_x\})\}$. By the Weak Stochastic Certainty Effect axiom, this implies that $\{\lambda p + (1-\lambda)q\} = \text{supp}_\rho(\{\lambda p + (1-\lambda)q, \lambda p + (1-\lambda)\delta_y\})$ for every $\lambda \in [0,1)$. Since $\succsim$ is a Deliberate Stochastic Choice representation of ρ, this, in turn, implies that $p \succsim \lambda p + (1-\lambda)q \succsim \lambda p + (1-\lambda)\delta_y \succ \lambda p + (1-\lambda)\delta_z$, for every $\lambda \in [0,1)$. This can happen only if $\{p\} = \text{supp}_\rho(\{p, \delta_z\})$, which implies that $V(p) \geq z$. Since x, y and z were arbitrarily chosen, we conclude that $V(p) \geq V(q)$. ‖

We now need the following claims:

Claim 3. *The function V is continuous.*

Proof of Claim. Pick a convergent sequence $(p^m) \in \Delta^\infty$. Now pick any convergent subsequence, $V(p^{m_k})$, of $V(p^m)$ and let $x^k = V(p^{m_k})$, for every k, $x = \lim V(p^{m_k})$, and $p = \lim p^m$. If $x = w$, then it is clear that $V(p) \geq x$, so suppose that $x > w$. Pick $\delta > 0$ such that $x - \delta > w$ and fix $\varepsilon \in (0, \delta)$. For k large enough, we have that $x^k > x - \varepsilon > w$ and, therefore, $\{p^{m_k}\} = \text{supp}_\rho(\{p^{m_k}, \delta_{x-\varepsilon}\})$. By the continuity axiom, this implies that $\{p\} = \text{supp}_\rho(\{p, \delta_{x-\delta}\})$, which implies that $V(p) \geq x - \delta$. Since δ was arbitrarily chosen, we conclude that $V(p) \geq x$. If $x = b$, then it is clear that $V(p) \leq x$, so suppose that $x < b$. Pick $\delta > 0$ such that $x + \delta < b$ and fix $\varepsilon \in (0, \delta)$. For k large enough, we have that $x^k < x + \varepsilon < b$ and, therefore, $\delta_{x+\varepsilon} \in \text{supp}_\rho(\{p^{m_k}, \delta_{x+\varepsilon}\})$. By the continuity axiom, this implies that $\{\delta_{x+\delta}\} = \text{supp}_\rho(\{p, \delta_{x+\delta}\})$, which implies that $V(p) \leq x + \delta$. Since δ was arbitrarily chosen, we conclude that $V(p) \leq x$. This shows that $V(p) = x = \lim V(p^{m_k})$. We have just shown that every convergent subsequence of $(V(p^m))$ converges to $V(p)$. Since $(V(p^m))$ is bounded, this implies that $V(p^m) \to V(p)$. ‖

[25] We note that, since ρ admits a Deliberate Stochastic Choice representation, $w \in D_p$ for every $p \in \Delta$, so that V is well-defined.

Claim 4. *For every $p, q \in \Delta$ and $x \in [w, b]$, if $V(p) \geq V(\delta_x)$, then $V(\lambda p + (1-\lambda)q) \geq V(\lambda\delta_x + (1-\lambda)q)$ for every $\lambda \in [0, 1]$.*

Proof of Claim. Fix $p \in \Delta$ and $x \in [w, b]$ with $V(p) \geq V(\delta_x) = x$. Fix $\lambda \in (0, 1)$ and $q \in \Delta$. Suppose first that $x = w$. If $V(\lambda\delta_x + (1-\lambda)q) = w$ or $p = \delta_x$, we have nothing to prove, so suppose that $V(\lambda\delta_x + (1-\lambda)q) > w$, $p \neq \delta_x$ and fix $z \in (w, V(\lambda\delta_x + (1-\lambda)q))$. By the definition of V, we know that $\{\lambda\delta_x + (1-\lambda)q\} = \mathrm{supp}_\rho(\{\lambda\delta_x + (1-\lambda)q, \delta_y\})$ for any $y \in (z, V(\lambda\delta_x + (1-\lambda)q))$. By the Weak Stochastic Certainty Effect axiom, this implies that $\{\gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\lambda\delta_x + (1-\lambda)q)\} = \mathrm{supp}_\rho(\{\gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\lambda\delta_x + (1-\lambda)q), \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)\delta_z\})\}$ for every $\gamma \in [0, 1)$. Since ρ admits a Deliberate Stochastic Choice representation, this can happen only if $\lambda p + (1-\lambda)q \succ \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\lambda\delta_x + (1-\lambda)q) \succsim \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)\delta_z$ for every $\gamma \in [0, 1)$.[26] This now implies that $\{\lambda p + (1-\lambda)q\} = \mathrm{supp}_\rho(\{\lambda p + (1-\lambda)q, \delta_z\})$ and we learn that $V(\lambda p + (1-\lambda)q) \geq z$. Since z was arbitrarily chosen, we conclude that $V(\lambda p + (1-\lambda)q) \geq V(\lambda\delta_x + (1-\lambda)q)$. Now suppose that $x > w$ and fix any $y \in (w, x)$. By the definition of V, we know that $\{p\} = \mathrm{supp}_\rho(\{p, \delta_{y'}\})$ for any $y' \in (y, x)$. The Weak Stochastic Certainty Effect axiom now implies that $\{\lambda p + (1-\lambda)q\} = \mathrm{supp}_\rho(\{\lambda p + (1-\lambda)q, \lambda\delta_y + (1-\lambda)q\})$, which implies that $\lambda p + (1-\lambda)q \succsim \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\lambda\delta_y + (1-\lambda)q)$ for every $\gamma \in [0, 1]$. If $V(\lambda\delta_y + (1-\lambda)q) = w$, then it is clear that $V(\lambda p + (1-\lambda)q) \geq V(\lambda\delta_y + (1-\lambda)q)$. Otherwise, pick any $z \in [w, V(\lambda\delta_y + (1-\lambda)q))$. For any $z' \in (z, V(\lambda\delta_y + (1-\lambda)q))$, we have $\{\lambda\delta_y + (1-\lambda)q\} = \mathrm{supp}_\rho(\{\lambda\delta_y + (1-\lambda)q, \delta_{z'}\})$. By Weak Stochastic Certainty Effect, this implies that $\{\gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\{\lambda\delta_y + (1-\lambda)q\})\} = \mathrm{supp}_\rho(\{\gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\{\lambda\delta_y + (1-\lambda)q\}), \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)\delta_{\hat{z}}\})$ for every $\gamma \in [0, 1]$ and every $\hat{z} \in (z, z')$. But then we have

$$\begin{aligned} \lambda p + (1-\lambda)q &\succsim \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\lambda\delta_y + (1-\lambda)q) \\ &\succsim \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)\delta_{\hat{z}} \\ &\succ \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)\delta_z, \end{aligned}$$

for every $\gamma \in [0, 1)$. This can happen only if $\{\lambda p + (1-\lambda)q\} = \mathrm{supp}_\rho(\{\lambda p + (1-\lambda)q, \delta_z\})$. Since z was arbitrarily chosen, we conclude that $V(\lambda p + (1-\lambda)q) \geq V(\lambda\delta_y + (1-\lambda)q)$. Since y was arbitrarily chosen and V is continuous, this now implies that $V(\lambda p + (1-\lambda)q) \geq V(\lambda\delta_x + (1-\lambda)q)$. ‖

Claim 5. *The function V satisfies risk aversion, in the sense that if q is a mean preserving spread of p, then $V(p) \geq V(q)$.*

Proof of Claim. Suppose that q is a mean preserving spread of p. If $V(q) = w$, then we have nothing to prove, so suppose that $V(q) > w$, and pick $y \in (w, V(q))$. By the definition of V, we know that $\{q\} = \mathrm{supp}_\rho(\{q, \delta_x\})$ for every $x \in (y, V(q))$. Now the risk aversion axiom implies that $\{p\} = \mathrm{supp}_\rho(\{p, \delta_y\})$ and, consequently, $V(p) \geq y$. Since y was arbitrarily chosen, we conclude that $V(p) \geq V(q)$. ‖

[26] Recall that we are assuming that $x = w$, for now.

Claim 6. *For every $x, y \in [w, b]$, if $x > y$, then $V(\lambda\delta_x + (1-\lambda)\delta_w) > V(\lambda\delta_y + (1-\lambda)\delta_w)$ for every $\lambda \in (0, 1]$.*

Proof of Claim. Fix $z \in (y, x)$. Let $\hat{x} = V(\lambda\delta_x + (1-\lambda)\delta_w)$. If $\hat{x} = b$, then the fact that ρ admits a Deliberate Stochastic Choice representation implies that $\{\delta_{\hat{x}}\} = \text{supp}_\rho(\{\lambda\delta_z + (1-\lambda)\delta_w, \delta_{\hat{x}}\})$. Otherwise, $\delta_{\hat{x}+\varepsilon} \in \text{supp}_\rho(\{\lambda\delta_x + (1-\lambda)\delta_w, \delta_{\hat{x}+\varepsilon}\})$ for every $\varepsilon > 0$ with $\hat{x} + \varepsilon \leq b$, and the continuity axiom implies that $\{\delta_{\hat{x}}\} = \text{supp}_\rho(\{\lambda\delta_z + (1-\lambda)\delta_w, \delta_{\hat{x}}\})$. Now let $\hat{y} = V(\lambda\delta_y + (1-\lambda)\delta_w)$. If $\hat{y} = w$, then the fact that ρ admits a Deliberate Stochastic Choice representation implies that $\{\lambda\delta_z + (1-\lambda)\delta_w\} = \text{supp}_\rho(\{\lambda\delta_z + (1-\lambda)\delta_w, \delta_{\hat{y}}\})$. Otherwise, $\{\lambda\delta_y + (1-\lambda)\delta_w\} = \text{supp}_\rho(\{\lambda\delta_y + (1-\lambda)\delta_w, \delta_{\hat{y}-\varepsilon}\})$ for every $\varepsilon > 0$ with $\hat{y} - \varepsilon \geq w$, and the continuity axiom implies that $\{\lambda\delta_z + (1-\lambda)\delta_w\} = \text{supp}_\rho(\{\lambda\delta_z + (1-\lambda)\delta_w, \delta_{\hat{y}}\})$. Since $\{\delta_{\hat{x}}\} = \text{supp}_\rho(\{\lambda\delta_z + (1-\lambda)\delta_w, \delta_{\hat{x}}\})$, but $\{\lambda\delta_z + (1-\lambda)\delta_w\} = \text{supp}_\rho(\{\lambda\delta_z + (1-\lambda)\delta_w, \delta_{\hat{y}}\})$, we conclude that $\delta_{\hat{x}} \succsim \lambda\delta_z + (1-\lambda)\delta_w \succsim \delta_{\hat{y}}$ and $\hat{x} \neq \hat{y}$. This can happen only if $V(\lambda\delta_x + (1-\lambda)\delta_w) = \hat{x} > \hat{y} = V(\lambda\delta_y + (1-\lambda)\delta_w)$. ‖

Now let $\hat{\succsim}$ be the relation induced by V. That is, let $\hat{\succsim}$ be defined by $p \hat{\succsim} q$ if and only if $V(p) \geq V(q)$. The claims above show that $\hat{\succsim}$ satisfies all the axioms in the statement of Theorem 4. This implies that there exists a compact set $\mathcal{W} \subseteq C([w, b])$ such that every function $v \in \mathcal{W}$ is strictly increasing and concave and, for every $p, q \in \Delta$,

$$p \hat{\succsim} q \iff \min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_p(v)) \geq \min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_q(v)).$$

By Claim 2, this gives us the desired representation.

Conversely, suppose now that ρ can be represented by a compact set $\mathcal{W} \subseteq C([w, b])$, as in the statement of the theorem. Define $V : \Delta \to \mathbb{R}$ by

$$V(p) = \min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_p(v)),$$

for every $p \in \Delta$. We can easily check that V is continuous and satisfies risk aversion, in the sense that if p and q in Δ are such that q is a mean preserving spread of p, then $V(p) \geq V(q)$. It is also easy to see that if p and q in Δ are such that p strictly first order stochastically dominates q, then $V(p) > V(q)$. Finally, we can check that if $p \in \Delta$ and $x \in [w, b]$ are such that $V(p) \geq$ (resp. $>$) x, then $V(\lambda p + (1 - \lambda q)) \geq$ (resp. $>$) $V(\lambda\delta_x + (1-\lambda)q)$ for every $q \in \Delta$ and $\lambda \in (0, 1]$. Similarly, if $x \geq$ (resp. $>$) $V(p)$, then $x \geq$ (resp. $>$) $V(\lambda p + (1-\lambda)\delta_x)$ for all $\lambda \in (0, 1]$.

The preorder $\succsim$ represented by V is a Deliberate Stochastic Choice representation of ρ. By Theorem 1, we know that ρ satisfies Rational Mixing. Now suppose that $p \in \Delta$ and $x \in (w, b]$ are such that $\{p\} = \text{supp}_\rho(\{p, \delta_x\})$. Fix $\lambda \in (0, 1]$, $q \in \Delta$ and $y \in [w, x)$. The fact that $\{p\} = \text{supp}_\rho(\{p, \delta_x\})$ implies that $\mathbb{E}_p(v) \geq x$ for every $v \in \mathcal{W}$. Consequently, $\mathbb{E}_{\lambda p + (1-\lambda)q}(v) > \mathbb{E}_{\lambda\delta_y + (1-\lambda)q}(v)$ for every $v \in \mathcal{W}$. In fact, this implies that $\mathbb{E}_{\lambda p + (1-\lambda)q}(v) > \mathbb{E}_{\gamma(\lambda\delta_y + (1-\lambda)q) + (1-\gamma)(\lambda p + (1-\lambda)q)}(v)$ for every $\gamma \in (0, 1]$. Since $\mathcal{W}$ is compact, this implies that $V(\lambda p + (1-\lambda)q) > V(\gamma(\lambda\delta_y + (1-\lambda)q) +$

$(1-\gamma)(\lambda p+(1-\lambda)q))$ for every $\gamma \in (0,1]$. This now implies that $\{\lambda p+(1-\lambda)q\} = \text{supp}_\rho(\{\lambda p+(1-\lambda)q, \lambda\delta_y+(1-\lambda)q\})$. We conclude that ρ satisfies Weak Stochastic Certainty Effect.

Now consider two convergent sequences $(p^m) \in \Delta^\infty$ and $(x^m) \in [w,b]^\infty$. Let $p = \lim p^m$ and $x = \lim x^m$. Pick $y \in [w,x)$ and let $q \in \Delta$ be such that q strictly first order stochastically dominates p. If $\{p^m\} = \text{supp}_\rho(\{p^m, \delta_x\})$ for every m, then $V(p^m) \geq x$ for every m. This implies that $V(p) \geq x > y$. By the representation of V, this implies that, for every $\lambda \in [0,1)$, $V(p) \geq V(\lambda p+(1-\lambda)\delta_x) > V(\lambda p+(1-\lambda)\delta_y)$. This can happen only if $\{p\} = \text{supp}_\rho(\{p, \delta_y\})$. Suppose now that $\delta_y \in \text{supp}_\rho(\{p^m, \delta_y\})$. This implies that $y \geq V(p^m)$ for every m. Since V is continuous, we learn that $x > y \geq V(p)$. From the representation of ρ, we know that this can happen only if $\{\delta_x\} = \text{supp}_\rho(\{p, \delta_x\})$. Now suppose that $\{p\} = \text{supp}_\rho(\{p, \delta_{x^m}\})$ for every m. This implies that $V(p) \geq x^m$ for every m. Since V agrees with strict first order stochastic dominance, this implies that $V(q) > V(p) \geq x$. But then we must have $\{q\} = \text{supp}_\rho(\{q, \delta_x\})$. Finally, suppose that $\delta_{x^m} \in \text{supp}_\rho(\{q, \delta_{x^m}\})$ for every m. Again, this implies that $x^m \geq V(q)$ for every m. Since V agrees with strict first order stochastic dominance, we learn that $x \geq V(q) > V(p)$. This now gives us that $\{\delta_x\} = \text{supp}_\rho(\{p, \delta_x\})$. This shows that ρ satisfies the continuity axiom.

Finally, suppose that the lotteries p and q in Δ are such that q is a mean preserving spread of p and $x \in (w,b]$ is such that $\{q\} = \text{supp}_\rho(\{q, \delta_x\})$. This implies that $V(q) \geq x$. Since V satisfies risk aversion, this implies that $V(p) > y$ for every $y \in [w,x)$. Consequently, we have that $V(p) = V(\lambda p+(1-\lambda)p) > V(\lambda\delta_y+(1-\lambda)p)$ for every $\lambda \in (0,1]$ and $y \in [w,x)$. This can happen only if $\{p\} = \text{supp}_\rho(\{p, \delta_y\})$ for every $y \in [w,x)$. This shows that ρ satisfies the risk aversion axiom. ■

Proof of Proposition 1. We say that a binary relation $\succsim$ has a point of strict convexity if there exist $p, q \in \Delta$ and $\lambda \in (0,1)$ such that

$$\lambda p+(1-\lambda)q \succ p, q.$$

Also, given a Cautious Stochastic Choice representation $\mathcal{W}$ of ρ, define $\succsim$ as the preference induced by

$$V(p) = \min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_p(v)).$$

Notice that $\succsim$ is continuous and strictly preserves first order stochastic dominance, so that it is a continuous Deliberate Stochastic Choice representation of ρ. By Theorem 2, we know that ρ is a non-degenerate stochastic choice function if and only if $\succsim$ has a point of strict convexity.

We now turn to prove the proposition. For part 1, suppose that ρ is a non-degenerate stochastic choice function. Then, it follows that $\succsim$ must admit a point of strict convexity, and thus must violate the independence axiom. This implies $|\mathcal{W}| > 1$. Moreover, there must exist $p, q \in \Delta$, $\lambda \in (0,1)$ such that $\lambda q+(1-\lambda)p \succ p, q$. Say without loss of generality that $p \succsim q$. Take $z \in [w,b]$ such that $\delta_z \sim p \succsim q$. Notice that by the model we must have $\lambda\delta_z + (1-\lambda)p \sim p$. This means that we have $\delta_z \succsim q$ and $\lambda q+(1-\lambda)p \succ \lambda\delta_z+(1-\lambda)p$. Take $x, y \in [w, b]$ such that $x > y > z$ but such that both are close enough to z that we have $\delta_y \succ q$ and $\lambda q+(1-\lambda)p \succ \lambda\delta_x+(1-\lambda)p$. Then, notice that we must have $\{\delta_y\} = \text{supp}_\rho(\{\delta_y, q\})$ by the representation. Moreover, we must also have that $\{\lambda\delta_x + (1-\lambda)p\} \neq \text{supp}_\rho(\{\lambda q + (1-\lambda)p, \lambda\delta_x + (1-\lambda)p\})$. This proves that ρ exhibits Certainty Bias.

Consider now part 2. Suppose that ρ exhibits Certainty Bias and $|\mathcal{W}| < \infty$. Then, there exist $x, y \in [w, b]$ with $x > y$, $A \in \mathcal{A}$, $r \in \Delta$ and $\lambda \in (0, 1)$ such that $\{\delta_y\} = \text{supp}_\rho(A \cup \{\delta_y\})$ and $\{\lambda\delta_x + (1-\lambda)r\} \neq \text{supp}_\rho(\lambda(A \cup \{\delta_x\} + (1-\lambda)r)$. Thus $\delta_y \succsim q$ for all $q \in co(A \cup \{\delta_y\})$ and there exists $p \in co(A)$ and $\gamma \in (0, 1]$ such that $\gamma(\lambda p + (1-\lambda)r) + (1-\gamma)(\lambda\delta_x + (1-\lambda)r) \succsim \lambda\delta_x + (1-\lambda)r$. Since $x > y$, we must then have $\delta_x \succ p$, which means that $\succsim$ violates the independence axiom. This implies $|\mathcal{W}| > 1$. Since we know that $\succsim$ is convex, then either $\succsim$ satisfies the Betweenness axiom, or it must admit a point of strict convexity. By Cerreia-Vioglio et al. (2018, Proposition 3) and since $|\mathcal{W}| < \infty$ and $\succsim$ violates independence, $\succsim$ violates Betweenness. This implies that $\succsim$ must admit a point of strict convexity. By the previous claim, ρ must be a non-degenerate stochastic choice function. ■

Proof of Proposition 2. Let $\succsim$ be the continuous Deliberate Stochastic Choice representation of a stochastic choice function ρ and suppose that ρ satisfies Linearity. Fix any pair of lotteries p and q such that $p \succ q$. We claim that we must have that $\rho(\{p, q\})(p) = 1$. To see that, let λ^* be the minimum value in $[0, 1]$ such that $\lambda^* p + (1-\lambda^*)q \succsim \lambda p + (1-\lambda)q$ for every $\lambda \in [0, 1]$. Since $\succsim$ represents ρ, we must have that $\rho(\{\lambda^* p + (1-\lambda^*)q, q\})(\lambda^* p + (1-\lambda^*)q) = 1$. By Linearity, we get that $\rho(\{p, q\})(p) = 1$. Now fix any pair of lotteries p and q in Δ with $p \succsim q$. Since $\succsim$ strictly preserves first order stochastic dominance, we have that, for any $\lambda \in (0, 1)$, $\lambda p + (1-\lambda)\delta_b \succ \lambda q + (1-\lambda)\delta_w$. By what we have just learned, this implies that $\rho(\{\lambda p + (1-\lambda)\delta_b, \lambda q + (1-\lambda)\delta_w\})(\lambda p + (1-\lambda)\delta_b) = 1$. By Linearity, we have that, for any $\alpha \in (0, 1)$ and $r \in \Delta$, $\rho(\{\alpha(\lambda p + (1-\lambda)\delta_b) + (1-\alpha)r, \alpha(\lambda q + (1-\lambda)\delta_w) + (1-\alpha)r\})(\alpha(\lambda p + (1-\lambda)\delta_b) + (1-\alpha)r) = 1$. Since $\succsim$ represents ρ, this implies that $\alpha(\lambda p + (1-\lambda)\delta_b) + (1-\alpha)r \succsim \alpha(\lambda q + (1-\lambda)\delta_w) + (1-\alpha)r$. Since this is true for any $\lambda \in (0, 1)$, continuity of $\succsim$ implies that $\alpha p + (1-\alpha)r \succsim \alpha q + (1-\alpha)r$. We have just shown that, for any $p, q \in \Delta$ with $p \succsim q$, we have $\alpha p + (1-\alpha)r \succsim \alpha q + (1-\alpha)r$, for every $\alpha \in (0, 1)$ and $r \in \Delta$. Since $\succsim$ is continuous, it is well-known that this implies that it admits an expected-utility representation. That is, there exists a continuous function $u : [w, b] \to \mathbb{R}$ such that, for every pair of lotteries p and q in Δ, $p \succsim q$ if, and only if, $\mathbb{E}_p(u) \geq \mathbb{E}_q(u)$. The proposition is now an immediate consequence of this observation. ■

References

Afriat, S. N. (1967): "The Construction of Utility Functions from Expenditure Data," *International Economic Review*, 8.

Agranov, M. and P. Ortoleva (2017): "Stochastic Choice and Preferences for Randomization," *Journal of Political Economy*, 125, 40–68.

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

Aliprantis, C. D. and O. Burkinshaw (1998): *Principles of Real Analysis*, Academic Press, New York.

Andreoni, J. and C. Sprenger (2010): "Certain and Uncertain Utility: The Allais Paradox and Five Decision Theory Phenomena," Mimeo University of California San Diego.

Ballinger, T. P. and N. T. Wilcox (1997): "Decisions, error and heterogeneity," *The Economic Journal*, 107, 1090–1105.

Barberá, S. and P. K. Pattanaik (1986): "Falmagne and the rationalizability of stochastic choices in terms of random orderings," *Econometrica*, 707–715.

Becker, G. M., M. H. DeGroot, and J. Marschak (1963): "An experimental study of some stochastic models for wagers," *Behavioral Science*, 8, 199–202.

Bogacz, R., E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen (2006): "The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks." *Psychological review*, 113, 700.

Busemeyer, J. R. and J. T. Townsend (1993): "Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment." *Psychological review*, 100, 432–459.

Camerer, C. F. (1989): "Does the Basketball Market Believe in the 'Hot Hand'?" *American Economic Review*, 79, pp. 1257–1261.

Camerer, C. F. and T.-H. Ho (1994): "Violations of the betweenness axiom and nonlinearity in probability," *Journal of risk and uncertainty*, 8, 167–196.

Cerreia-Vioglio, S. (2009): "Maxmin Expected Utility on a Subjective State Space: Convex Preferences under Risk," Mimeo, Bocconi University.

Cerreia-Vioglio, S., D. Dillenberger, and P. Ortoleva (2015): "Cautious Expected Utility and the Certainty Effect," *Econometrica*, 83, 693–728.

——— (2018): "An Explicit Representation for Disappointment Aversion and Other Betweenness Preferences," Mimeo, Princeton University.

Chambers, C. P. and F. Echenique (2016): *Revealed preference theory*, Econometric Society Monograph; Cambridge University Press.

Chew, S. H. (1989): "Axiomatic utility theories with the betweenness property," *Annals of Operations Research*, 19, 273–298.

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

Cohen, M. A. (1980): "Random utility systems—the infinite case," *Journal of Mathematical Psychology*, 22, 1–23.

Dean, M. and P. Ortoleva (2015): "Is it All Connected? A Testing Ground for Unified Theories of Behavioral Economics Phenomena," Mimeo, Columbia University.

——— (2017): "Allais, Ellsberg, and Preferences for Hedging," *Theoretical Economics*, 12, 377–424.

Dekel, E. (1986): "An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom," *Journal of Economic Theory*, 40, 304–318.

Dillenberger, D. (2010): "Preferences for One-Shot Resolution of Uncertainty and Allais-Type Behavior," *Econometrica*, 78, 1973–2004.

Dwenger, N., D. Kübler, and G. Weizsäcker (2016): "Flipping a Coin: Evidence from Laboratory and Field," Mimeo Humboldt-Universität zu Berlin.

Epper, T. and H. Fehr-Duda (2015): "The missing link: Unifying risk taking and time discounting," Mimeo, University of Zurich.

Falmagne, J.-C. (1978): "A Representation Theorem for Random Finite Scale Systems," *Journal of Mathematical Psychology*, 18, 52–72.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015): "Stochastic Choice and Revealed Perturbed Utility," *Econometrica*, 83, 2371–2409.

Fudenberg, D. and T. Strzalecki (2015): "Recursive Logit with Choice Aversion," *Econometrica*, 83, 651–691.

Gul, F. (1991): "A theory of disappointment aversion," *Econometrica*, 59, 667–686.

Gul, F., P. Natenzon, and W. Pesendorfer (2014): "Random choice as behavioral optimization," *Econometrica*, 82, 1873–1912.

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

Harless, D. W. and C. F. Camerer (1994): "The predictive utility of generalized expected utility theories," *Econometrica*, 1251–1289.

Harsanyi, J. C. (1973): "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points," *International Journal of Game Theory*, 2, 1–23.

Hey, J. (2001): "Does repetition improve consistency?" *Experimental economics*, 4, 5–54.

Hey, J. and C. Orme (1994): "Investigating generalizations of expected utility theory using experimental data," *Econometrica*, 1291–1326.

Johnson, E. J. and R. Ratcliff (2013): "Computational and Process Models of Decision Making in Psychology and Behavioral Economics," in *Neuroeconomics: Decision Making and the Brain*, ed. by P. W. Glimcher and E. Fehr, Academic Press, 35–48.

Kahneman, D. and A. Tversky (1979): "Prospect theory: an analysis of choice under risk," *Econometrica*, 47, 263–291.

Kircher, P., S. Ludwig, and A. Sandroni (2013): "On the Difference between Social and Private Goods," *The BE Journal of Theoretical Economics*, 13.

Loomes, G. and R. Sugden (1995): "Incorporating a stochastic element into decision theories," *European Economic Review*, 39, 641–648.

Luce, R. D. (1958): "A probabilistic theory of utility," *Econometrica: Journal of the Econometric Society*, 193–224.

——— (1959): *Individual choice behavior: A theoretical analysis*, Wiley (New York).

Machina, M. J. (1985): "Stochastic choice functions generated from deterministic preferences over lotteries," *The Economic Journal*, 95, 575–594.

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

Marley, A. (1997): "Probabilistic choice as a consequence of nonlinear (sub) optimization," *Journal of Mathematical Psychology*, 41, 382–391.

McFadden, D. (2006): "Revealed stochastic preference: a synthesis," in *Rationality and Equilibrium*, ed. by C. Aliprantis, R. Matzkin, D. McFadden, J. Moore, and N. Yannelis, Springer Berlin Heidelberg, vol. 26 of *Studies in Economic Theory*, 1–20.

McFadden, D. and M. K. Richter (1991): "Stochastic Rationality and Revealed Stochastic Preference," in *Preferences, Uncertainty, and Rationality*, ed. by J. Chipman, D. McFadden, and K. Richter, Westview Press, 161–186.

Nishimura, H., E. A. Ok, and J. K.-H. Quah (2017): "A comprehensive approach to revealed preference theory," *American Economic Review*, 107, 1239–63.

Ortoleva, P. (2010): "Status Quo Bias, Multiple Priors and Uncertainty Aversion," *Games and Economic Behavior*, 69, 411–424.

Quiggin, J. (1982): "A theory of anticipated utility," *Journal of Economic Behavior & Organization*, 3, 323–343.

Ratcliff, R. and G. McKoon (2008): "The diffusion decision model: theory and data for two-choice decision tasks," *Neural computation*, 20, 873–922.

Ratcliff, R. and P. L. Smith (2004): "A comparison of sequential sampling models for two-choice reaction time." *Psychological review*, 111, 333.

Regenwetter, M., J. Dana, and C. P. Davis-Stober (2011): "Transitivity of preferences," *Psychol Rev*, 118, 42–56.

Regenwetter, M. and C. P. Davis-Stober (2012): "Behavioral variability of choices versus structural inconsistency of preferences," *Psychol Rev*, 119, 408–16.

Rockafellar (1970): *Convex Analysis*, Princeton University Press, Princeton.

Sopher, B. and J. M. Narramore (2000): "Stochastic choice and consistency in decision making under risk: An experimental study," *Theory and Decision*, 48, 323–350.

Starmer, C. and R. Sugden (1989): "Probability and juxtaposition effects: An experimental investigation of the common ratio effect," *Journal of Risk and Uncertainty*, 2, 159–178.

Stoye, J. (2015): "Choice theory when agents can randomize," *Journal of Economic Theory*, 155, 131–151.

Swait, J. and A. Marley (2013): "Probabilistic choice (models) as a result of balancing multiple goals," *Journal of Mathematical Psychology.*

Thurstone, L. L. (1927): "A law of comparative judgment." *Psychological review*, 34, 273.

Tversky, A. (1969): "Intransitivity of preferences." *Psychological review*, 76, 31.

Woodford, M. (2014): "Stochastic Choice: An Optimizing Neuroeconomic Model," *American Economic Review: Papers & Proceedings*, 104, 495–500.

Wu, G. and R. Gonzalez (1996): "Curvature of the probability weighting function," *Management science*, 42, 1676–1690.