Deliberately Stochastic*

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

First version: February 2012
This version: December 4, 2016
Latest version available at: www.columbia.edu/~po2205/papers/DeliberatelyStochastic.pdf

Abstract
We study stochastic choice as the outcome of *deliberate* randomization. First, we present a general model in which the agent has a preference relation over lotteries and the stochastic choice is the optimal mix among available options. The model is characterized by a rationality-type axiom on the stochastic choice. Second, we study the special case in which the stochastic choice reflects underlying preferences that belong to the Cautious Expected Utility class (Cerreia-Vioglio et al., 2015a). This is characterized by behavior that is consistent with a weak form of certainty bias.

JEL: *D80, D81*

Keywords: *Stochastic Choice, Random Utility, Hedging, Cautious Expected Utility, Convex Preferences*

---

*We thank Eddie Dekel, Kfir Eliaz, Faruk Gul, Peter Klibanoff, Efe Ok, Wolfgang Pesendorfer, Marciano Siniscalchi and the audience at various seminars and conferences for their useful comments. Ortoleva gratefully acknowledges the financial support of NSF grant SES-1156091, and and Riella gratefully acknowledges the financial support of CNPq of Brazil, grant no. 304560/2015-4.
[†]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, Columbia University, email: `pietro.ortoleva@columbia.edu`.
[¶]ENAE, Getulio Vargas Foundation, 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, also 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 is thus seen as 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, who may *choose* to report different answers. For example, Machina (1985) notes that this is precisely what the agent may wish to do if her preferences over lotteries or acts are *convex*, that is, exhibit affinity towards randomization between equally good options. Crucially, convexity is a typical implication of many existing models of decision making under risk and uncertainty developed to address the Allais or Ellsberg paradoxes. Such desire to randomize is thus an implication of many existing models. Convexity of preferences also has experimental support (Becker et al., 1963; Sopher and Narramore, 2000). A small theoretical lit-

[1]To avoid confusion, note that 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 (2015).

[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 reviews, Ratcliff and Smith (2004), Bogacz et al. (2006), Johnson and Ratcliff (2013).

erature suggested also other, different reasons for stochastic choice to be deliberate: Marley (1997) and Swait and Marley (2013) follow lines similar to Machina (1985), Dwenger et al. (2013) suggest it may be due to a desire to minimize regret, and Fudenberg et al. (2015), connect it to uncertain taste shocks. In Section 4 we discuss these papers in detail.

Recent experimental evidence supports the interpretation of stochastic choice as deliberate. Agranov and Ortoleva (2015) 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. Dwenger et al. (2013) 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 matters, chosen randomly. They find that a significant fraction of students report inconsistent rankings, even when there are no strategic reasons to do so.[5]

In this paper we develop axiomatically two 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 desire may be a rational reaction if the underlying preferences are 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. First, we prove a very general representation theorem: we show that a rationality-type condition on stochastic choice, reminiscent of the acyclicity of the revealed preference relation used in choice from limited datasets, guarantees that it can be represented as if the agent were choosing the *optimal* mixing over the existing options given an underlying complete preference relation over the final monetary lotteries. In this model the stochasticity has a purely *instrumental* value for the agent: she doesn't 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 final distribution they induce, rather than as compound lotteries.

The model above imposes only minimal requirements on the underlying preferences. To obtain more structure, we note that a desire to mix must derive from violations of Expected Utility. Additional structure will thus result from imposing regularities on when such violations can occur. Since one of the most robust evidence of such violations is the certainty bias, as captured by the Allais paradox, we posit that violations cannot occur in ways that are explicitly in the *opposite* direction

[5]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 chose to randomize. Rubinstein (2002) documents a deliberate desire to report “diversified” answers even when this leads to strictly dominated choices; other recent studies document “false” diversification (Chen and Corter, 2006; Eliaz and Fréchette, 2008).

– strictly certainty "averse" –, at least in the extreme case in which the stochastic choice is degenerate.[6] This, together with continuity and risk aversion, characterizes the special case of the general model in which the underlying preferences are represented by the Cautious Expected Utility model of Cerreia-Vioglio et al. (2015a): the agent has a *set* of utility functions over outcomes, and evaluates each lottery p by displaying a cautious behavior. She computes 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. Intuitively, the desire to mix between options emerges from the agent's subjective uncertainty of how to evaluate lotteries; she may benefit from 'hedging' in a similar way to how an ambiguity averse agent may benefit from hedging between two acts. We call this the *Cautious Stochastic Choice* model.

Note that both in the Cautious Stochastic Choice model and in models of Random Utility there are multiple utilities being considered. With Random Utility, one is randomly picked, whereas in the former all 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. In Section 4 we show that to this difference in interpretation corresponds a difference in behavior: the only case in which stochastic choice can be modeled with both Random Utility and our Cautious Stochastic Choice model is when the Random Utility has only one utility and stochastic choice takes place only when the agent is fully indifferent.

We conclude by emphasizing how our model entail a strong form of "rationality" on the part of the agent: she acts as if she foresees the consequences of each possible randomization and chooses the best one. This may be only partially realistic – other forces may be at play. One may thus see our results as providing an "extreme-rationality" model of stochastic choice as the outcome of deliberate randomization of subjects whose underlying preferences may violate Expected Utility.[7]

The remainder of the paper is organized as follows. Section 2 presents the general Deliberate Stochastic Choice model. Section 3 presents the Cautious Stochastic Choice model. Section 4 discusses the related literature. The proofs appear in the Appendix.

---

[6]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 is degenerate; it posits no restrictions when the stochastic choice is not degenerate. It is thus conceptually much weaker.

[7]Violations of Expected Utility for risk may still be considered rational, and may take place in ways that are disjunct from, for example, failure to reduce compound lotteries. See for example Segal (1992). (Empirically, Dean and Ortoleva (2015) show how these behaviors are uncorrelated.)

2 A general model of deliberately stochastic choice

2.1 Framework and Foundations

Let $[w, b] \subset \mathbb{R}$ be an 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 set of finite subsets of Δ. For any $A \in \mathcal{A}$, $\text{co}(A)$ 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 and $A \in \mathcal{A}$, $\text{supp}_\rho(A)$ denotes the support of $\rho(A)$.

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,

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

Note that 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 final distribution it induces.[8]

We can now discuss our first axiom. Our goal is to capture behaviorally an agent who is *deliberately* choosing her stochastic choice following an underlying preference 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 $co(A)$. 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 in A_2. Since we don't 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.

Axiom 1 (Rational Hedging)**.** *For every $k \in \mathbb{N} \backslash \{1\}$ 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 $q \not>_{FOSD} \overline{\rho(A_1)}$.

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

Rational Hedging 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 (Chambers and Echenique, 2015). 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 Hedging 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 final distribution over lotteries, not valuable *per se*. This implies a form of reduction of compound lotteries, which we will maintain throughout.

2.2 Deliberate Random 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) $\succeq$ over Δ such that:

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

Theorem 1. *A stochastic choice ρ satisfies Rational Hedging if, and only if, ρ admits a Deliberate Stochastic Choice representation.*

A Deliberate Stochastic Choice model captures a decision maker who has preferences $\succeq$ over monetary lotteries and chooses deliberately the randomization that generates the optimal mixture among existing options. This is most prominent when $\succeq$ 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.

Note that 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 $\succeq$ follows Expected Utility, or satisfies Betweenness;[9] or even aversion to randomization, e.g., if $\succeq$ are Rank Dependent Expected Utility (RDU) preferences with pessimistic distortions: in these cases the agent has no desire to mix and the stochastic choice is degenerate.[10]

[9] 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).

[10] 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.

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 discountinuities.[11]

Our analysis above assumes that we observe the stochastic choice of the agent 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, i.e., the stochastic choice only from sets in $\mathcal{B} \subset \mathcal{A}$. The following observation shows that the Rational Hedging axiom is necessary and sufficient in *any* dataset. (The proof is omitted as it follows exactly the same steps as the proof of Theorem 1, proved for the general case.)

Observation 1. Consider the set $\mathcal{B} \subset \mathcal{A}$ and denote by $\rho_{\mathcal{B}}$ the restriction of ρ on $\mathcal{B}$. Then, $\rho_{\mathcal{B}}$ satisfies Rational Hedging if and only if $\rho_{\mathcal{B}}$ admits a Deliberate Stochastic Choice representation.

Finally, note that the preference relation in a Deliberate Stochastic Choice model need not admit a utility representation. For this we need a form of continuity. To posit it, 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 a randomization ($p = \overline{\rho(A)}$), from a set A where q could have also been chosen ($q \in co(A)$). Denote by $tran(R)$ the transitive closure of R.

Axiom 2 (Continuity)**.** *$tran(R)$ is closed.*

Proposition 1. *If a stochastic choice ρ satisfies Rational Hedging and Continuity, then it admits a Deliberate Stochastic Choice representation where $\succeq$ that can be represented by a continuous utility function.*

3 The Cautious Stochastic Choice model

Theorem 1 has the benefits and drawbacks of generality: it captures stochastic choice as the deliberate desire to report a stochastic answer, with few further assumptions; but it puts only minimal restrictions on preferences over lotteries, thus providing limited predictive power. We now turn to a special case in which we give a specific functional form representation to the underlying preferences $\succeq$.

[11]While with choice correspondences the continuity of the underlying preference relation implies continuity of the choice correspondence, 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.

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 Bias, 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 Bias). *For any $p, q \in \Delta(X)$, $x, y \in [w, b]$ 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, as defined, for example, in 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 Bias is related to Negative Certainty Independence, or NCI (Dillenberger, 2010; Cerreia-Vioglio et al., 2015a), which is imposed on preferences over lotteries and requires that if p is preferred to δ_x, then a mixture of p and q should be weakly preferred to a mixture 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 Bias is instead imposed on stochastic choice and, crucially, only on the extreme case in which the stochastic choice is *degenerate*. It imposes no restrictions on behavior for sets where the stochastic choice 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 discountinuities of ρ due to this, as we have discussed above (see ft. 11).

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:*

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

Next, we impose Risk Aversion – noting that this is imposed here for purely technical reasons, as it is behaviorally distinct from deliberate stochasticity.[12] 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 any $p, q \in \Delta$, if q is a mean preserving spread of p and $\{q\} = \operatorname{supp}_\rho(\{q, \delta_x\})$ for some $x \in [w,b]$, then $\{p\} = \operatorname{supp}_\rho(\{p, \delta_y\})$ for every $y \in [w,x)$.*

3.1 The Cautious Stochastic Choice model

We are now ready to introduce the second main 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 2. A stochastic choice 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

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

Theorem 2. *Let ρ be a stochastic choice on Δ. The following statements are equivalent:*

[12] The reason why it is needed is again related to the fact that we do not restrict how indifference is broken. Without risk aversion, we would obtain a similar representation but without the requirement that all functions are concave and that the set of utilities is compact. The latter is needed to guarantee that the underlying preferences are strictly increasing with respect to first order stochastic dominance, which in turn is essential to identify indifferences.

1. *The stochastic choice ρ satisfies Rational Hedging, Weak Stochastic Certainty Bias, Continuity* and Risk Aversion;*

2. *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 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. (2015a).[13] This is derived here by imposing Weak Stochastic Certainty Bias, which is reminiscent of NCI, which in turn characterizes the Cautious Expected Utility model (see Cerreia-Vioglio et al. 2015a). However, we have seen that Weak Stochastic Certainty Bias does not apply to the (unobserved) underlying preference, but constrains behavior only in extreme situations where the stochastic choice 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 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. 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 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.

In terms of uniqueness, the representation of $\succeq$ is unique.

[13] It is a special case in that $\mathcal{W}$ is compact and all utilities are concave. (The latter follows from risk aversion, Cerreia-Vioglio et al. 2015a.)

Proposition 2. *Consider two Cautious Stochastic Choice representations $\mathcal{W}$ and $\mathcal{W}'$ of some stochastic choice function ρ. Then for every $q \in \Delta$*

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

Less straightforward is the uniqueness of $\mathcal{W}$. This is very similar to the uniqueness properties in Cerreia-Vioglio et al. (2015a), 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 closed convex hull of $\mathcal{W}$, $\overline{co}(\mathcal{W})$, 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 $u \in \mathcal{W}$. 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 3. *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 $\widehat{\mathcal{W}} \subseteq \mathcal{W}$.*

3.2 Stocastic Choice and Certianty Bias

In the Cautious Stochastic Choice model the underlying preference of the agent follows the 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, and Cerreia-Vioglio et al. (2015a) show that it is again the multiplicity of utilities and the caution that generate this behavior. 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.

First, we say that an agent exhibits a non-degenerate stochastic choice if stochasticity is present not only when the agent is indifferent: 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 3. We say that a stochastic choice ρ is *non-degenerate* if there exists $p, q \in \Delta$ with $|\text{supp}_\rho(\{p, q\})| \neq 1$ and $\lambda \in (0, 1)$ such that

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

We can also define a strict version of Certainty Bias, where at least once a strict advantage is given to certainty.

Definition 4. We say that a stochastic choice ρ exhibits *Certainty Bias* 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 4. *Consider a stochastic choice function ρ that admits a Cautious Stochastic Choice representation. Then the following holds:*

1. *if ρ is a non-degenerate stochastic choice, then ρ must exhibit Certainty Bias and all Cautious Stochastic Choice representations of ρ must have $|\mathcal{W}| > 1$;*

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

The proposition above shows a connection between Certainty Bias and stochasticity of choice. Note that this relation is not an if and only if: if $\mathcal{W}$ 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 exhibit Certainty Bias (for example, 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 Certainty Bias and stochasticity suggested above finds experimental support in Agranov and Ortoleva (2015), where the stochasticity of answers is correlated with the tendency to exhibit the Allais paradox.

4 Related Literature

This paper is related to various strands of the literature. First, it is connected with 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.[14] Dwenger et al. (2013) 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

[14] 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 Hedging 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.)

choice is stochastic rather than deterministic. Fudenberg et al. (2015) study different variational representations of a stochastic choice function. In their model, the agent chooses the probability distribution that maximizes Expected Utility minus the (psychological) cost of choosing. Similarly to our interpretation, they also argue that their model corresponds to a form of ambiguity-averse preferences for an agent who is uncertain about her true utility. The main difference between the two models is that we study a domain of menus of lotteries, and interpret, in the spirit of Machina (1985), stochastic choice as a manifestation of nonlinear preferences over lotteries. In Fudenberg et al. (2015), on the other hand, the domain is menus of final outcomes, and their interpretation of deliberate randomization by subjects relies on the existence of uncertain taste shocks, which are chosen by nature in a way that gives the agent incentive to randomize. This implies substantial differences both between the axioms and the models.

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 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 (Dekel 1986; Chew 1989), 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. (2015a). As previously mentioned, our analysis here differs as we can't observe the preferences, and thus impose a postulate reminiscent of NCI but only for the extreme case in which the stochastic choice 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,[15] 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, 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 most 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

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

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 (2015) and Nishimura et al. (2015).

We conclude by comparing our model to the class 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 x in a choice problem A, the probability of choosing x from A, $\rho(A)(x)$, equals the probability of drawing a utility function u such that x maximizes u in A.[16] 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).

It is well-known that a stochastic choice function that admits a Random Utility representation satisfies the following properties:

Monotonicity. For every $A, B \in \mathcal{A}$ and $x \in A$, if $A \subseteq B$ then $\rho(B)(x) \leq \rho(A)(x)$.

Revealed Dominance. $\rho(\{x, y\})(y) = 0$ implies $\rho(A)(y) = 0$ whenever $x \in A$.

We show that adding these two properties to the Cautious Stochastic Choice model implies that, in the absence of indifferences, the individual *never* randomizes.

Proposition 5. *Suppose that ρ is a stochastic choice function that admits a Cautious Stochastic Choice representation $\mathcal{W}$ and, in addition, satisfies Monotonicity and Revealed Dominance. Define the function $V : \Delta \to \mathbb{R}$ by*

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

Then, for every $A \in \mathcal{A}$

$$\operatorname{supp}_\rho(A) \subseteq \arg\max_{p \in A} \ V(p).$$

The proposition above shows that the intersection of the models of Cautious Stochastic Choice and Random Utility is a *degenerate* random utility, in which only *one* utility, V, is used. In fact, while the representation above is similar to the general Cautious Stochastic Choice model, the support includes only the lotteries that maximize V in the original set A, not in its convex hull. This means that randomizations happen only to break ties when the agent is indifferent.

Gul and Pesendorfer (2006) axiomatizes 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:

[16] 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. (Gul and Pesendorfer (2006) calls this property regularity). Second, to impose a tie-breaking rule independent of the choice problem. Our results below hold in both cases.

Linearity. For every $A \in \mathcal{A}$ and $p \in A$, $\rho(A)(p) = \rho(\lambda A + (1-\lambda)q)(\lambda p + (1-\lambda)q)$ for every $q \in \Delta$ and $\lambda \in (0,1)$.

We now show that if ρ is a Cautious Stochastic Choice model that in addition satisfies Monotonicity, Revealed Dominance and Linearity, then ρ is a *degenerate* Random Expected Utility model, i.e., again a model with only one utility. Formally:

Proposition 6. *Suppose that ρ is a stochastic choice function that admits a Cautious Stochastic Choice representation $\mathcal{W}$ and, in addition, satisfies Monotonicity, Revealed Dominance and Linearity. Then there exists a continuous function $u : [w,b] \to \mathbb{R}$ such that, for any choice problem A,*

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

The results above are summarized in Figure 1.

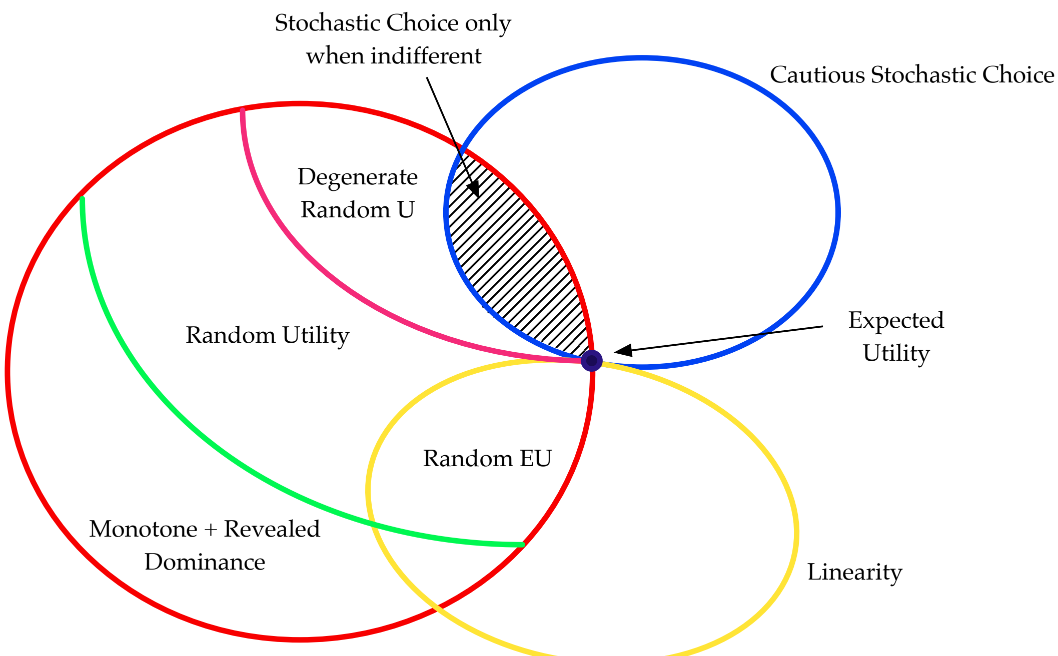

Figure 1: Relation with models in the literature

Appendix A: Preliminary Results

In this section we present a result that extends the analysis in Cerreia-Vioglio et al. (2015a). For that, as we did in the main text, let $[w, b]$ be a closed 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 $\succsim \subseteq \Delta \times \Delta$. We will impose the following postulates on $\succsim$:

Axiom 6 (Weak Order)**.** *The relation $\succsim$ is complete and transitive.*

Axiom 7 (Continuity)**.** *For every $q \in \Delta$, the sets $\{p \in \Delta : p \succsim q\}$ and $\{p \in \Delta : q \succsim p\}$ are closed.*

Axiom 8 (Negative Certainty Independence (NCI))**.** *For every $p, q \in \Delta$, $x \in [w, b]$ and $\lambda \in [0, 1]$,*

$$p \succsim \delta_x \text{ implies } \lambda p + (1-\lambda)q \succsim \lambda\delta_x + (1-\lambda)q.$$

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

$$x > y \text{ implies } \lambda\delta_x + (1-\lambda)\delta_w \succ \lambda\delta_y + (1-\lambda)\delta_w.$$

Axiom 10 (Risk Aversion)**.** *For every pair of lotteries $p, q \in \Delta$, if q is a mean preserving spread of p, then $p \succsim q$.*

We can now enunciate the following theorem:

Theorem 3. *Let $\succsim$ be a binary relation on Δ. The following statements are equivalent:*

1. *The relation $\succsim$ satisfies Weak Order, Continuity, Monotonicity, Negative Certainty Independence, and Risk Aversion;*
2. *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 $p, q \in \Delta$,*

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

Proof of Theorem 3.

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 4 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. It is trivial that (ii) implies (i). For the other direction, consider $\{v_n\}_{n\in\mathbb{N}} \subseteq \mathcal{W}$. Observe that, by construction, $\{v_n\}_{n\in\mathbb{N}}$ is uniformly bounded. By assumption, there exists $\{v_{n_k}\}_{k\in\mathbb{N}} \subseteq \{v_n\}_{n\in\mathbb{N}}$ 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, pag. 65) 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 4. *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 $\lambda \in (0, 1]$

$$x > y \Longrightarrow V(\lambda\delta_x + (1-\lambda)\delta_w) > V(\lambda\delta_y + (1-\lambda)\delta_w) \tag{2}$$

then $\mathcal{W}$ is relatively compact with respect to the topology of sequential pointwise convergence restricted to $\mathcal{U}$.

Proof of Theorem. Let's first show that, for each $\varepsilon > 0$, there exists $\delta > 0$ *such that, for each $v \in \mathcal{W}$,*

$$v(w + \delta) < \varepsilon.$$

To see that, assume that there exists $\bar{\varepsilon} > 0$ such that for each $\delta > 0$ 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

$$\lambda_k v_k(b) + (1-\lambda_k) v_k(w) = \lambda_k = v_k\left(w + \frac{1}{k}\right) \geq \bar{\varepsilon} > 0. \tag{3}$$

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 (3) 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 (2).

Now, consider $\{v_n\}_{n\in\mathbb{N}} \subseteq \mathcal{W}$. Observe that, by construction, $\{v_n\}_{n\in\mathbb{N}}$ is uniformly bounded. By (Rockafellar, 1970, Theorem 10.9), there exists $\{v_{n_k}\}_{k\in\mathbb{N}} \subseteq \{v_n\}_{n\in\mathbb{N}}$ 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 $\{v_{n_k}\}_{k\in\mathbb{N}} \subseteq \mathcal{W}$, 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 $\bar{v} \in \mathcal{W}$, 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 the initial claim, 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 $\{v_{n_k}\}_{k\in\mathbb{N}}$, 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 $\{\lambda_k\}_{k\in\mathbb{N}} \subseteq [0,1)$ 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 $\{v_{n_k}\}_{k\in\mathbb{N}}$, 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. ‖

We now complete the proof of Theorem 3.

(i) implies (ii). By the results in Cerreia-Vioglio et al. (2015a), there exists $\mathcal{W} \subseteq \mathcal{U}_{nor}$ such that the function $V : \Delta \to \mathbb{R}$, defined by

$$V(p) = \inf_{v\in\mathcal{W}} v^{-1}(\mathbb{E}_p v) \qquad \forall p \in \Delta, \tag{4}$$

is a continuous utility function for $\succsim$. Moreover, $\mathcal{W}$ is closed under the topology of sequential pointwise convergence restricted to $\mathcal{U}$ and each $v \in \mathcal{W}$ is concave. 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 4, it follows that $\mathcal{W}$ is in fact compact under the topology of sequential pointwise convergence restricted to $\mathcal{U}$. This yields compactness in the topology of sequential pointwise convergence. 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 (4) 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 the results in Cerreia-Vioglio et al. (2015a), $\succsim$ satisfies Weak Order, Continuity, Weak Monotonicity, 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 $\succcurlyeq$ satisfies Strict First Order Stochastic Dominance and so, in particular, Monotonicity.

Appendix B: Proof of the Results in the Text

Proof of Theorem 1 and Proposition 1.

It is clear that if ρ admits a Deliberate Stochastic Choice representation, then ρ satisfies Rational Hedging. Suppose, thus, that ρ satisfies Rational Hedging and define the binary relation R the same way it is defined in the main text. 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 Hedging 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 Hedging 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 $\succeq$ of $tran(R)$. By what we have just seen, $\succeq$ is also an extension of the first order stochastic dominance relation. Moreover, by definition, we have that $\overline{\rho(A)} R q$ for every $q \in co(A)$, for every $A \in \mathcal{A}$. Consequently, we have $\overline{\rho(A)} \succeq q$ for every $q \in co(A)$, for every $A \in \mathcal{A}$. This proves Theorem 1.

Suppose now that, in addition, ρ also satisfies Contintuity. Then $tran(R)$ is a continuous preorder. By Levin's Theorem, there exists a continuous function $u : \Delta \to \mathbb{R}$ such that $p \; tran(R) \; q$ implies $u(p) \geq u(q)$, with strict inequality whenever it is not true that $q \; tran(R) \; p$. We can use this function u and the preference it induces to finish the proof of Proposition 1.

Proof of Theorem 2.

Suppose first that ρ satisfies all the axioms in the statement of the theorem. By Theorem 1, ρ admits a Deliberate stochastic choice Representation $\succeq$. 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 $p \succeq \lambda p + (1 - \lambda)\delta_x$ for every $\lambda \in [0, 1]$. Since $\succeq$ 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

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

We note that 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 $p \in \Delta$.[17] 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 a $q \in co(A)$. 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 Bias 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 $\succeq$ is a Deliberate Stochastic Choice Representation of ρ, this, in turn, implies that $p \succeq \lambda p + (1 - \lambda)q \succeq \lambda p + (1 - \lambda)\delta_y \succ \lambda p + (1 - \lambda)\delta_z$, 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,

[17] 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.

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 $V(p^m) \to V(p)$. ‖

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 $\{\lambda\delta_x + (1-\lambda)q\} = \text{supp}_\rho(\{\lambda\delta_x + (1-\lambda)q, \delta_y\})$ for any $y \in (z, V(\lambda\delta_x + (1-\lambda)q))$. By the Degenerate Hedging axiom, this implies that $\{\gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\lambda\delta_x + (1-\lambda)q)\} = \text{supp}_\rho(\{\gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\lambda\delta_x + (1-\lambda)q), \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)\delta_z\})\}$ for every $\gamma \in [0,1)$. Since ρ admits a Deliberate Stochastic Choice Representation, this can happen only if $\lambda p + (1-\lambda)q \succ \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\lambda\delta_x + (1-\lambda)q) \succeq \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)\delta_z$ for every $\gamma \in [0,1)$.[18] This now implies that $\{\lambda p + (1-\lambda)q\} = \text{supp}_\rho(\{\lambda p + (1-\lambda)q, \delta_z\})$ 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 $\{p\} = \text{supp}_\rho(\{p, \delta_{y'}\})$ for any $y' \in (y, x)$. The Weak Stochastic Certainty Bias axiom now implies that $\{\lambda p + (1-\lambda)q\} = \text{supp}_\rho(\{\lambda p + (1-\lambda)q, \lambda\delta_y + (1-\lambda)q\})$, which implies that $\lambda p + (1-\lambda)q \succeq \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\lambda\delta_y + (1-\lambda)q)$ 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 $\{\lambda\delta_y + (1-\lambda)q\} = \text{supp}_\rho(\{\lambda\delta_y + (1-\lambda)q, \delta_{z'}\})$. By Weak Stochastic Certainty Bias, this implies that $\{\gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\{\lambda\delta_y + (1-\lambda)q\})\} = \text{supp}_\rho(\{\gamma(\lambda p + (1-\lambda)q) + (1-\gamma)(\{\lambda\delta_y + (1-\lambda)q\}), \gamma(\lambda p + (1-\lambda)q) + (1-\gamma)\delta_{\hat{z}}\})$ for every $\gamma \in [0,1]$ and every $\hat{z} \in (z, z')$. But then we have

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

for every $\gamma \in [0,1)$. This can happen only if $\{\lambda p + (1-\lambda)q\} = \text{supp}_\rho(\{\lambda p + (1-\lambda)q, \delta_z\})$. Since z was arbitrarily chosen, we conclude that $V(\lambda p + (1-\lambda)q) \geq$

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

$V(\lambda\delta_y + (1-\lambda)q)$. 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 $\{q\} = \text{supp}_\rho(\{q, \delta_x\})$ for every $x \in (y, V(q))$. Now the risk aversion axiom implies that $\{p\} = \text{supp}_\rho(\{p, \delta_y\})$ and, consequently, $V(p) \geq y$. Since y was arbitrarily chosen, we conclude that $V(p) \geq V(q)$. ‖

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 $\delta_{\hat{x}} \succeq \lambda\delta_z + (1-\lambda)\delta_w \succeq \delta_{\hat{y}}$ 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 $\succsim$ be the relation induced by V. That is, let $\succsim$ be defined by $p \succsim p$ if, and only if, $V(p) \geq V(q)$. The claims above show that $\succsim$ satisfies all the axioms in the statement of Theorem 3. 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 $p, q \in \Delta$,

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

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 stochasticaly dominates q, then $V(p) > V(q)$. Finally, we can check that if $p \in \Delta$ and $x \in [w,b]$ are such that $V(p) \geq^{(>)} x$, then $V(\lambda p + (1-\lambda q)) \geq^{(>)} V(\lambda\delta_x + (1-\lambda)q)$ for every $q \in \Delta$ and $\lambda \in (0,1]$. Similarly, if $x \geq^{(>)} V(p)$, then $x \geq^{(>)} V(\lambda p + (1-\lambda)\delta_x)$ for all $\lambda \in (0,1]$.

The preorder $\succeq$ represented by V is a Deliberate Stochastic Choice representation of ρ. By Theorem 1, we know that ρ satisfies Rational Hedging. 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 Bias.

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 stochasticaly 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 imples 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 2.

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 $q \in \Delta$,

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

To see that, first notice that it is clear from the fact that $\mathcal{W}$ is a Cautious Stochastic

Choice representation of ρ that $\{q\} = \text{supp}_\rho(\{q, \delta_y\})$ for every $y \in [w, b]$ with

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

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

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

This can happen only if

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

Proof of Proposition 3.

By Proposition 2, any two Cautious Stochastic Choice representations of the same stochastic choice function ρ are Cautious Expected Utility representations of the same binary relation $\succcurlyeq$. Now Theorem 2 in (Cerreia-Vioglio et al., 2015a) immediately implies that there exists a convex and normalized Cautious Stochastic Choice representation $\widehat{\mathcal{W}}$ of ρ such that $\widehat{\mathcal{W}} \subseteq \overline{co}(\mathcal{W})$ for any other normalized Cautious Stochastic Choice Representation of ρ.

Proof of Proposition 4.

We say that a binary relation $\succeq$ 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 $\succeq$ as the preference induced by

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

Notice that $\succeq$ is continuous and satisfies strict first order stochastic dominance.

We start by proving the following claim.

Claim 1. *ρ is a non-degenerate stochastic choice if and only if $\succeq$ has a point of strict convexity.*

Proof. Suppose that ρ is a non-degenerate stochastic choice. Then, there exist $p, q \in \Delta$ and $\lambda \in (0, 1)$ such that $p, q \in \text{supp}_\rho(\{p, q\})$ and

$$\lambda p + (1 - \lambda)\delta_w, q \in \text{supp}_\rho(\{\lambda p + (1 - \lambda)\delta_w, q\}),$$

and

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

Say without loss of generality that we have $q \succeq p$. By the representation, this means that there exist $\alpha \in (0,1)$ such that $\alpha(\lambda p + (1-\lambda)\delta_w) + (1-\alpha)q \succeq q$, hence $\alpha p + (1-\alpha)q \succ p, q$. Thus $\succeq$ has a point of strict convexity.

Conversely, suppose that there exist $p, q \in \Delta$, $\lambda \in (0,1)$ such that $\lambda p + (1-\lambda)q \succ p, q$. Then, we must have $p, q \in \text{supp}_\rho(\{p, q\})$. Moreover, by continuity, there must exist a $\gamma \in (0,1)$ such that $\lambda(\gamma\delta_w + (1-\gamma)p) + (1-\lambda)q \succ p, q$ and $\lambda p + (1-\lambda)(\gamma\delta_w + (1-\gamma)q \succ p, q$. Thus, ρ must be a non-degenerate stochastic choice. $\|$

We now turn to prove the proposition. For part 1, suppose that ρ is a non-degenerate stochastic choice function. Then, it follows that $\succeq$ 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 $p \succeq q$. Take $z \in [w, b]$ such that $\delta_z \sim p \succeq q$. Notice that by the model we must have $\lambda\delta_z + (1-\lambda)p \sim p$. This means that we have $\delta_z \succeq q$ 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 $\delta_y \succeq q$ for all $q \in co(A \cup \{\delta_y\})$ and there exists $p \in co(A)$ and $\gamma \in (0,1]$ such that $\gamma(\lambda p + (1-\lambda)r) + (1-\gamma)(\lambda\delta_x + (1-\lambda)r) \succeq \lambda\delta_x + (1-\lambda)r$. Since $x > y$, we must then have $\delta_x \succ p$, which means that $\succeq$ violates the independence axiom. This implies $|\mathcal{W}| > 1$. Since we know that $\succeq$ is convex, then either $\succeq$ satisfies the Betweenness axiom, or it must admit a point of strict convexity. But Cerreia-Vioglio et al. (2015b) proves that a Cautious Expected Utility model represents a preference relation that satisfies Betweenness only when the representation has an infinite set of utilities. Since $|\mathcal{W}| < \infty$, then $\succeq$ must admit a point of strict convexity. By the previous claim, ρ must be a non-degenerate stochastic choice.

Proof of Proposition 5.

Suppose that ρ admits a Cautious Stochastic Choice representation $\mathcal{W}$ and satisfies the other two properties mentioned in the statement of the proposition. Define $V : \Delta \to \mathbb{R}$ as

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

Fix a choice problem A and pick a lottery

$$q \notin \arg\max_{p \in A} V(p).$$

Fix

$$p^* \in \arg\max_{p \in A} V(p).$$

Pick $x \in [w, b]$ such that

$$V(p^*) > x > V(q).$$

By the representation of ρ, we have that $\text{supp}_\rho(\{p^*, \delta_x\}) = \{p^*\}$ and $\text{supp}_\rho(\{q, \delta_x\}) = \{\delta_x\}$. By Revealed Dominance, this implies that $\{q, \delta_x\} \cap \text{supp}_\rho(A \cup \{\delta_x\}) = \emptyset$. But then, if $q \in \text{supp}_\rho(A)$ Monotonicity would be violated for some alternative $p \in \text{supp}_\rho(A \cup \{\delta_x\})$. We conclude that $q \notin \text{supp}_\rho(A)$.

Proof of Proposition 6.

Define $V : \Delta \to \mathbb{R}$ by

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

for every $p \in \Delta$. By Proposition 5, for every pair of lotteries p and q in Δ, if $V(p) > V(q)$, then $\{p\} = \text{supp}_\rho(\{p, q\})$. Fix $\lambda \in (0, 1)$ and $r \in \Delta$. By Linearity, we must have $\{\lambda p + (1-\lambda) r\} = \text{supp}_\rho(\{\lambda p + (1-\lambda) r, \lambda q + (1-\lambda) r\})$. Applying Proposition 5 again, we learn that $V(\lambda p + (1-\lambda) r) \geq V(\lambda q + (1-\lambda) r)$. We can use a similar reasoning to show that we must have $V(p) \geq V(q)$ whenever $V(\lambda p + (1-\lambda) r) > V(\lambda q + (1-\lambda) r)$ for some $\lambda \in (0, 1)$ and $r \in \Delta$. We can now use the fact that V is continuous and agrees with strict first order stochastic dominance to show that the preference relation represented by V satisfies the independence axiom. Since it is also continuous, it can be represented, in the expected utility sense, by a continuous function $u : [w, b] \to \mathbb{R}$. Applying Proposition 5, we obtain the desired conclusion.

References

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

Agranov, M. and P. Ortoleva (2015): "Stochastic Choice and Preferences for Randomization," *Journal of Political Economy, forthcoming.*

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.

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 (2015a): "Cautious Expected Utility and the Certainty Effect," *Econometrica*, 83, 693–728.

——— (2015b): "An explicit representation for preferences under risk," Mimeo, Columbia University.

Chambers, C. P. and F. Echenique (2015): *Revealed preference theory*, Cambridge University Press Econometric Society Monographs, forthcoming.

Chen, Y.-J. and J. E. Corter (2006): "When mixed options are preferred in multiple-trial decisions," *Journal of Behavioral Decision Making*, 19, 17–42.

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.

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 (2013): "Flipping a Coin: Theory and Evidence," Mimeo Humboldt-Universität zu Berlin.

Eliaz, K. and G. Fréchette (2008): "Don't Put All Your Eggs in One Basket!: An Experimental Study of False Diversification," Mimeo New York University.

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 (forthcoming)*.

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. (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 (2015): "A Unified Approach to Revealed Preference Theory: The Case of Rational Choice," Mimeo New York University.

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.

Rubinstein, A. (2002): "Irrational diversification in multiple decision problems," *European Economic Review*, 46, 1369–1378.

Segal, U. (1992): "The independence axiom versus the reduction axiom: Must we have both?" in *Utility Theories: Measurements and Applications*, Springer, 165–183.

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.