

Deliberately Stochastic*

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

First version: February 2012

This version: May 25, 2017

Abstract

We study stochastic choice as the outcome of *deliberate* randomization. After first deriving a general representation of a stochastic choice function with such property, we proceed to characterize a model in which the agent has preferences over lotteries that belong to the Cautious Expected Utility class (Cerreia-Vioglio et al., 2015a), and the stochastic choice is the optimal mix among available options. This model links stochasticity of choice and the phenomenon of Certainty Bias, with both behaviors stemming from the same source: multiple utilities and caution. We show that this model is behaviorally distinct from models of Random Utility, as it typically violates the property of Regularity, shared by all of them.

JEL: *D80, D81*

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

*We thank Eddie Dekel, Kfir Eliaz, Mira Frick, Faruk Gul, Ryota Iijima, Peter Klibanoff, 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.

[†]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, 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 show how stochastic choice can be seen as stemming from the same source as known violations of Expected Utility, like the Certainty Bias; and 3) to identify whether, and how, such models of deliberate randomization can be behaviorally distinguished from existing models of stochastic choice.

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

¹To 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.

²The 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).

³Thurstone (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).

⁴Models 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 4 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 matters, chosen randomly. 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 a desire may be a rational reaction if the underlying preferences over lotteries are convex; in particular, we aim to show how both stochasticity of choice and violations of Expected Utility in line with the well-known Allais paradoxes may be seen as stemming from the *same* source. 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.

We begin our analysis with 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

⁵In 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.

⁶Kircher 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.

can be represented as if the agent were choosing the *optimal* mixing over the existing options given an underlying complete and transitive preference relation over the final monetary lotteries. 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, rather than as compound lotteries.

The very general representation 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 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 – strictly certainty “averse” – at least in the extreme case in which the stochastic choice function is degenerate.⁷ The main result of the paper is to show that this postulate, together 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. (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 in the Cautious Stochastic Choice 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.

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

⁷Our 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.

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

Our last set of results relates the Cautious Stochastic Choice model with existing models of stochastic choice. We begin by considering a well-known property of stochastic choice, widely used in the literature: Regularity. It posits that the probability of choosing p from a set cannot decrease if we remove elements from it. We show that adding Regularity to our model implies that the agent acts as if she had one utility and randomized only in the cases of indifference – in all other cases the stochastic choice must be degenerate.⁹ Thus, a strict desire to hedge implies violations of Regularity. 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.

That the Cautious Stochastic Choice model is inconsistent with Regularity has direct implications on its relation with 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 of a degenerate random utility (only one utility), where randomization occurs only in the case of indifference. Thus, the conceptual difference highlighted above 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.

We conclude the introduction by emphasizing that our agent acts *as if* she foresees the final consequences of each possible randomization over lotteries and chooses the best one; a model which may be only partially realistic. One may thus see our results as providing a ‘rational’ model of stochastic choice as the outcome of deliberate randomization of subjects whose only deviation from standard behavior is possible violations of Expected Utility – all other possible deviations are ruled out.¹⁰

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

⁹In particular, we show that adding Regularity implies that the underlying preference satisfies a property called Betweenness (Dekel, 1986; Chew, 1989), which amounts to having linear indifference curves and thus gives the agent no incentive to randomize. See Section 4 for a discussion.

¹⁰In particular, our agents also do not fail to reduce compound lotteries, a property that has been both theoretically (Segal, 1992) and empirically (Dean and Ortoleva, 2017) shown to be disjunct from violations of Expected Utility.

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 and Section 4 discusses the relation with existing models. All proofs appear in the Appendix.

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 collection of all finite and nonempty 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 , $A \in \mathcal{A}$, 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,

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

Note that by construction, the convex hull of a set A , $\text{co}(A)$, 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.¹¹

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 $\text{co}(A)$, 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 $\overline{\rho(A_2)} \in \text{co}(A_1)$. This means that the lottery chosen from A_2 could be obtained also from A_1 . Standard rationality posits that the ‘choice’ from

¹¹That is, by construction we must have $\text{co}(A) = \{p \in \Delta : p = \overline{\rho(A)} \text{ for some stochastic choice function } \rho\}$.

$A_1, \overline{\rho(A_1)}$, 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.

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

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

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

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 (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 $p, q \in \Delta$, $p \succ_{FOSD} q$ 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 monetary 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.

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 \succsim follows Expected Utility, or satisfies Betweenness;¹² 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 the stochastic choice function is degenerate.¹³

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.¹⁴

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 $p \succ_{FOSD} q$. Consider any $\mathcal{B} \subset \mathcal{A}$ such that $\{p, q\} \in \mathcal{B}$ whenever $p \succ_{FOSD} q$ 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 in a Deliberate Stochastic Choice model need not admit a utility representation. For this we need a form of continuity. Proposition 7 in Appendix B gives a sufficient condition for that: in words, we need to add the additional requirement that the binary relation R , defined as pRq if $p = \overline{\rho(A)}$ for some A such that $q \in co(A)$, has a transitive closure that is a closed set.

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;

¹²That 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).

¹³If we order the prizes in the support of a finite lottery p , with $x_1 < x_2 < \dots < x_n$, 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 $f : [0, 1] \rightarrow [0, 1]$ is strictly increasing and onto and $u : [w, b] \rightarrow \mathbb{R}$ is increasing. We say that distortions are pessimistic if f is convex, which implies *aversion* to randomization.

¹⁴While 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.

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 \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 2 (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, 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 Effect is related to Negative Certainty Independence, or NCI (Dillenberger, 2010; Cerreia-Vioglio et al., 2015a), 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 14).

Axiom 3 (Continuity). *Let $(p^m) \in \Delta^\infty$ and $(x^m) \in [w, b]^\infty$ be convergent sequences with $p^m \rightarrow p$ and $x^m \rightarrow x$. Let $y \in (w, x)$ and $q \in \Delta$ be such that $q \succ_{FOSD} p$. Then:*

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

Next, we impose Risk Aversion – noting that this is imposed here for purely technical reasons, as it is behaviorally distinct from deliberate stochasticity.¹⁵ 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 4 (Risk Aversion). *For each $p, q \in \Delta$, if q is a mean preserving spread of p and $\{q\} = \text{supp}_\rho(\{q, \delta_x\})$ for some $x \in (w, b]$, then $\{p\} = \text{supp}_\rho(\{p, \delta_y\})$ for each $y \in [w, x)$.*

3.1 Representation Theorem

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, $\mathbb{E}_p(v) = \int_{[w, b]} v dp$.

¹⁵This is needed again because 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.

Definition 2. 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

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

Theorem 2. 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. (2015a).¹⁶ 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.

¹⁶It is a special case in that \mathcal{W} is compact and all utilities are concave. (The latter follows from risk aversion.)

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.¹⁷ 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 relation \succsim induced by a Cautious Stochastic Choice representation is unique.

Proposition 1. *Consider two Cautious Stochastic Choice representations \mathcal{W} and \mathcal{W}' of some stochastic choice function ρ . Then, for each $q \in \Delta$*

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

Less straightforward is the uniqueness of \mathcal{W} . These are similar uniqueness properties as 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} , $\overline{\text{co}}(\mathcal{W})$, 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 2. *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 $\widehat{\mathcal{W}} \subseteq \mathcal{W}$.*

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:

¹⁷A formal comparison between these models appears in Section 4.

Proposition 3. *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} .

3.2 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.

First, we say that an agent exhibits a non-degenerate stochastic choice function 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 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

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

We can also define the case in which the 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 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 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 $|\mathcal{W}| < \infty$ and $cl(co(\mathcal{W})) = \widehat{\mathcal{W}}$, where $\widehat{\mathcal{W}}$ is a set as in Proposition 2, 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 canonical representation with finitely many utilities ($|\mathcal{W}| < \infty$ and $cl(co(\mathcal{W})) = \widehat{\mathcal{W}}$), 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 if, and only if, all representations must 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 any canonical representation of ρ ($cl(co(\mathcal{W})) = \widehat{\mathcal{W}}$) 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 be Certainty 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.

4 Relation with Models in the Literature

Regularity. In this section we compare the Cautious Stochastic Choice model with existing models of stochastic choice. To this end, it will be useful to first discuss the relation with a well-known property of stochastic choice, extensively used in the literature.

Axiom 5 (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 element 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 choice of any of the remaining ones. In other words, the attractiveness of an option should not depend on the availability of other ones.

The next proposition shows that postulating that the Cautious Stochastic Choice model satisfies Regularity completely eliminates any benefit from mixing.

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

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

Then the following is true:

1. *For each $A \in \mathcal{A}$,*

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

2. *For every $p, q \in \Delta$ and $\lambda \in (0, 1)$,*

$$V(p) \geq V(q) \implies V(p) \geq V(\lambda p + (1 - \lambda)q) \geq V(q).$$

The proposition above shows that adding Regularity to the Cautious Stochastic Choice representation leads to a model in which the agent exhibits stochastic behavior *only when she is indifferent*: in (1), the support of the stochastic choice function includes only the lotteries that maximize V in the original set A , and not in its convex hull. This implies that the agent never randomizes (except for indifference). This is also clear from (2), which shows that in this case the agent has no benefit from hedging: mixing between two lotteries can never yield a higher utility than the best one of them – in particular, the preference relation induced by V satisfies Betweenness (See footnote 12).¹⁸

Proposition 5 implies that in any Cautious Stochastic Choice model, if randomization happens for a genuine desire to mix and not due to indifference, or, equivalently, the agents' preferences violate Betweenness, then the stochastic choice *must* violate Regularity. This is a crucial feature of the model. To illustrate this point, consider the following example:

Example 1. Let $X = [0, 20]$ and define functions u and v from X into \mathbb{R} by

$$u(x) = \begin{cases} x & \text{if } x \leq 5 \\ \frac{1}{5}x + 4 & \text{if } x \geq 5 \end{cases}$$

and

$$v(x) = \begin{cases} 6x - 10 & \text{if } x \leq 2 \\ x & \text{if } x \geq 2 \end{cases}.$$

¹⁸In fact, statements 1 and 2 in Proposition 5 are equivalent to each other (even in the absence of Regularity), so for the Cautious Stochastic Choice model randomizing only in case of indifferences and the impossibility of getting better off by mixing are the same thing. We prove this observation in Proposition 8. We also show, in Proposition 9, that Regularity is compatible with any Betweenness preference. In other words, from a preferential point of view, Regularity does not impose any additional restrictions beyond Betweenness.

Consider the lotteries p and q such that $p(0) = 1/5$, $p(5) = 4/5$, $q(2) = 4/5$ and $q(12) = 1/5$. Note that $u^{-1}(\mathbb{E}_p(u)) = 4$, $v^{-1}(\mathbb{E}_p(v)) = 2$, $u^{-1}(\mathbb{E}_q(u)) = 72/25$ and $v^{-1}(\mathbb{E}_q(v)) = 4$. Note that if ρ is a stochastic choice function with a Cautious Stochastic Choice representation $\mathcal{W} = \{u, v\}$, then:

$$\text{supp}_\rho(\{p, \delta_3\}) = \{\delta_3\} \quad \text{and} \quad \text{supp}_\rho(\{q, \delta_3\}) = \{\delta_3\}.$$

However, we must also have $\text{supp}_\rho(\{p, q, \delta_3\}) = \{p, q\}$:¹⁹ the agent has a strict desire to mix between p and q . This implies two violations of Regularity: we have $0 = \rho(\{p, \delta_3\})(p) < \rho(\{p, q, \delta_3\})(p)$ as well as $0 = \rho(\{q, \delta_3\})(q) < \rho(\{p, q, \delta_3\})(q)$.

Intuitively, in the example above when *both* p and q are present, the agent has an incentive to choose *both*, mixing them. This allows her to hedge and obtain a higher utility. The two lotteries are, in a way, *complementary* to each other. But if p is removed, the agent can no longer hedge, and she is thus no longer interested in choosing q . This generates a violation of Regularity. 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 it provides as a whole. 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.

One crucial aspect of this argument is that Regularity prescribes that the choice probability of any element should not decrease if we remove *any* other element, including chosen ones. Our model would in general satisfy a weaker version that posits that choice probabilities do not decrease if we remove *unchosen* elements – violations of this may occur only for different ways of breaking ties.

Random Utility. Proposition 5 can be used to easily compare the Cautious Expected Utility 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.

¹⁹In particular, we have $\rho(\{p, q, \delta_3\})(p) = 14/39$ and $\rho(\{p, q, \delta_3\})(q) = 25/39$.

²⁰Stochastic 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.

(In models of Random Utility, there is no complementarity between the chosen elements.) But then, Proposition 5 above shows 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 Cautious Expected Utility.

Random Expected Utility. 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:

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

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

We now show that if ρ is a Cautious Stochastic Choice model that in addition satisfies Regularity 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 Regularity and Linearity. Then there exists a continuous function $u : [w, b] \rightarrow \mathbb{R}$ 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

²¹Machina (1985) suggests the following condition: if $A, A' \in \mathcal{A}$ are such that $co(A) \subset co(A')$ and $\rho(A') \in co(A)$, then $\rho(A') = \rho(A)$. (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.)

depends only on the choice probabilities. Formally, they axiomatize a stochastic choice function ρ such that, for each choice problem A ,

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

where $\Delta(A)$ is the set of probability measures on A , u is a von Neumann-Morgenstern utility function and $c : [0, 1] \rightarrow \mathbb{R} \cup \{\infty\}$ is strictly convex and C^1 in $(0, 1)$. They call this representation *Weak Additive Perturbed Utility* (Weak APU) representation.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.

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 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

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

²³An 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.

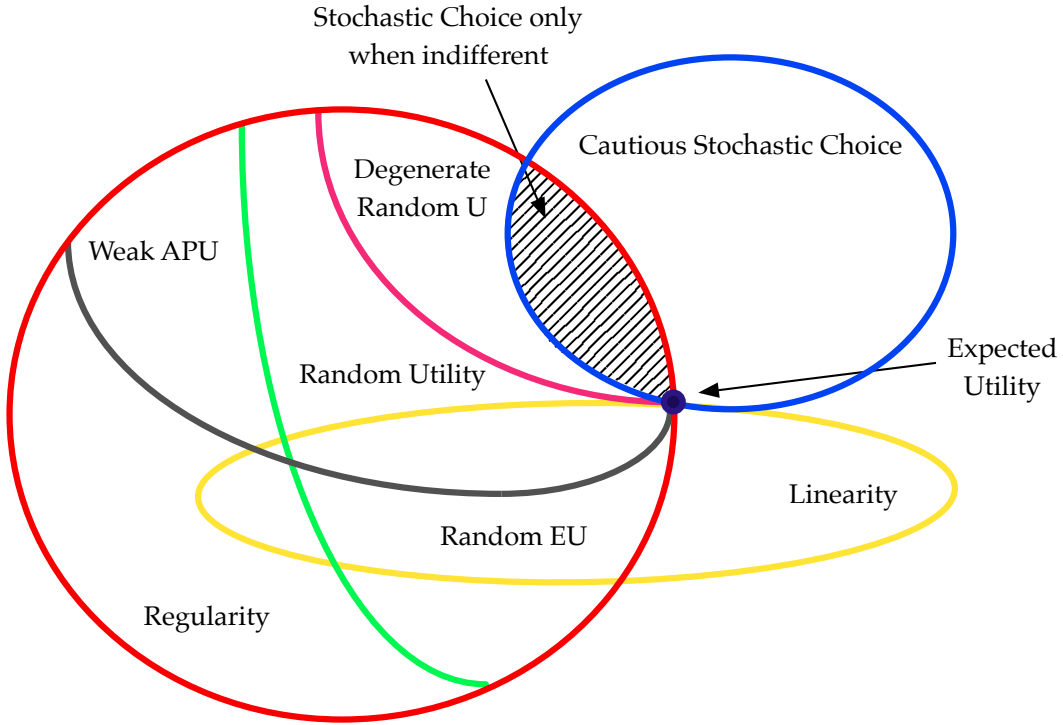


Figure 1: Relation with the Models in the Literature

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 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 most general representation theorem (Theorem 1) is related to the literature on revealed preferences on finite datasets. By randomizing

²⁴The 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.

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. (2015).

Appendix A: Preliminary Results

In this section we present a result that extends the analysis of Cerreia-Vioglio et al. (2015a) which will be instrumental to prove the results in this paper. 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 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 $\lambda \in [0, 1]$,*

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

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 3. *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 $p, q \in \Delta$,*

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

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 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. 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 $v_{n_k}(x) \rightarrow v(x)$ 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. \blacksquare

Theorem 4. Let $V : \Delta \rightarrow \mathbb{R}$ and $\mathcal{W} \subseteq \mathcal{U}_{nor}$ be such that

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

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

$$x > y \implies V(\lambda \delta_x + (1 - \lambda) \delta_w) > V(\lambda \delta_y + (1 - \lambda) \delta_w) \quad (3)$$

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

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

$$v(w + \delta) < \varepsilon. \quad (4)$$

By contradiction, 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 $v_k(w + \frac{1}{k}) \geq \bar{\varepsilon}$. Define $\lambda_k \in [0, 1]$ for each $k > \frac{1}{b-w}$ to be such that

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

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 $\lambda_k \rightarrow \lambda$. Notice that $\lambda \geq \bar{\varepsilon}$. Define $p = \lambda \delta_b + (1 - \lambda) \delta_w$. It is immediate to see that $p_k \rightarrow p$. By (5) and by definition of V , it follows that

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

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 $v_{n_k}(x) \rightarrow v(x)$ 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 $\bar{v} : [w, b] \rightarrow [0, 1]$ by

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

Since $v_{n_k} \in \mathcal{W} \subseteq \mathcal{U}_{nor}$ for every $k \in \mathbb{N}$, we have that $v_{n_k}(x) \rightarrow \bar{v}(x)$ 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 $1 \geq \limsup_{x \rightarrow b^-} \bar{v}(x) \geq \liminf_{x \rightarrow b^-} \bar{v}(x) \geq \lim_{x \rightarrow b^-} \frac{x-w}{b-w} = 1$, 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 $x \in [w, w + \delta)$

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

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 $\lambda_k v_{n_k}(b) + (1 - \lambda_k) v_{n_k}(w) = \lambda_k = v_{n_k}(x)$. Since \bar{v} is the pointwise limit of (v_{n_k}) , it follows that $\lambda_k \rightarrow 1$. Define $p_k = \lambda_k \delta_b + (1 - \lambda_k) \delta_w$ for all $k \in \mathbb{N}$. It is immediate to see that $p_k \rightarrow \delta_b$. 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. \blacksquare

Proof of Theorem 3 (i) implies (ii). By (Cerrea-Vioglio et al., 2015a, Theorem 2) and since Monotonicity implies Weak Monotonicity, there exists $\mathcal{W} \subseteq \mathcal{U}_{nor}$ such that the function $V : \Delta \rightarrow \mathbb{R}$, defined by

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

is a continuous utility function for \succsim . By (Cerrea-Vioglio et al., 2015a, 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 $\mathcal{W}_{\max-nor}$ in (Cerrea-Vioglio et al., 2015a, Proposition 4 and p. 720)). 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}_{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 $V : \Delta \rightarrow \mathbb{R}$ defined by

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

By hypothesis, V is well defined and it represents \succsim . Since \mathcal{W} is compact, we have that V is continuous. By (Cerrea-Vioglio et al., 2015a, 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 satisfies Strict First Order Stochastic Dominance and so, in particular, Monotonicity. \blacksquare

Appendix B: Additional Results

Continuous Deliberately Stochastic Choice Models

In this section we extend the results of Theorem 1 to the case in which the underlying preference relation admits 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 $\text{tran}(R)$ the transitive closure of R .

Axiom 12 (Continuity'). $\text{tran}(R)$ is closed.

Proposition 7. *If a stochastic choice function ρ satisfies Rational Hedging and Continuity', then it admits a Deliberate Stochastic Choice representation \succsim that can be represented by a continuous utility function.*

Proof of Proposition 7. Suppose that ρ also satisfies Continuity'. Then $\text{tran}(R)$ is a continuous preorder. Also, by the same argument used in the proof of Theorem 1, Rational Mixing implies that $\text{tran}(R)$ is an extension of the first order stochastic dominance relation. Moreover, by Levin's Theorem, there exists a continuous function $u : \Delta \rightarrow \mathbb{R}$ such that $p \text{ tran}(R) q$ implies $u(p) \geq u(q)$, with strict inequality whenever it is not true that $q \text{ tran}(R) p$. Now we can again proceed as in the proof of Theorem 1, using the preference relation the function u induces, to conclude the proof. \blacksquare

Cautious Stochastic Choice and Betweenness

In this section we show that statements 1 and 2 in Proposition 5 are equivalent for Cautious Stochastic Choice representations. Formally:

Proposition 8. *Suppose that ρ is a stochastic choice function that admits a Cautious Stochastic Choice representation \mathcal{W} and define the function $V : \Delta \rightarrow \mathbb{R}$ by*

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

The following statements are equivalent:

(i) *For each $A \in \mathcal{A}$,*

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

(ii) *The binary relation \succsim represented by V satisfies Betweenness. That is, for every $p, q \in \Delta$ and $\lambda \in (0, 1)$,*

$$p \succsim q \implies p \succsim \lambda p + (1 - \lambda)q \succsim q.$$

Proof of Proposition 8. Before we begin we first need the following claim.

Claim 1. *If V is defined as in (7) for some Cautious Stochastic Choice representation \mathcal{W} , then, for every pair of lotteries p and q in Δ , $V(p) > V(q)$ implies that $V(\lambda p + (1 - \lambda)q) > V(q)$ for every $\lambda \in (0, 1)$.*

Proof of Claim. Pick lotteries p and q such that $V(p) > V(q)$ and fix $\lambda \in (0, 1)$. Let $v \in \mathcal{W}$ be such that $v^{-1}(\mathbb{E}_{\lambda p + (1 - \lambda)q}(v)) = V(\lambda p + (1 - \lambda)q)$. Since $V(p) > V(q)$, we must have $\mathbb{E}_p(v) \geq v(V(p)) > v(V(q))$ and $\mathbb{E}_q(v) \geq v(V(q))$. But then $\mathbb{E}_{\lambda p + (1 - \lambda)q}(v) = \lambda \mathbb{E}_p(v) + (1 - \lambda) \mathbb{E}_q(v) > v(V(q))$, which implies that $V(\lambda p + (1 - \lambda)q) = v^{-1}(\mathbb{E}_{\lambda p + (1 - \lambda)q}(v)) > v^{-1}(v(V(q))) = V(q)$. \parallel

Now, let us show first that (ii) implies (i). By contradiction, assume that there exists $A \in \mathcal{A}$ such that $\text{supp}_\rho(A) \not\subseteq \arg \max_{p \in A} V(p)$. Since ρ is a stochastic choice function that admits a Cautious Stochastic Choice representation \mathcal{W} , we have that

$$\overline{\rho(A)} \in \arg \max_{p \in \text{co}(A)} V(p)$$

Recall that $\overline{\rho(A)} = \sum_{i=1}^n \lambda_i p_i$ where $n \in \mathbb{N}$, $\{p_i\}_{i=1}^n = \text{supp}_\rho(A) \subseteq A$, $\{\lambda_i\}_{i=1}^n \subseteq (0, 1]$, and $\sum_{i=1}^n \lambda_i = 1$. Without loss of generality, assume that $p_1 \succsim \dots \succsim p_n$. Since $\text{supp}_\rho(A) \not\subseteq \arg \max_{p \in A} V(p)$, we have that there exists $\bar{n} \in \{1, \dots, n\}$ such that $p_i \in \arg \max_{p \in A} V(p)$ for all $i \in \{1, \dots, \bar{n} - 1\}$ and $p_i \notin \arg \max_{p \in A} V(p)$ for all $i \in \{\bar{n}, \dots, n\}$.²⁵ We have two cases:

²⁵If $\bar{n} = 1$, we set $\{1, \dots, \bar{n} - 1\} = \emptyset$.

1. $\bar{n} = 1$. In this case, let $\hat{p} \in \Delta$ be such that $\hat{p} \in \operatorname{argmax}_{p \in A} V(p)$. By construction, note that $\hat{p} \succ p_i$ for all $i \in \{1, \dots, n\}$. Since \succsim satisfies Betweenness, we can conclude that

$$A \ni \hat{p} \succ p_1 \succsim \sum_{i=1}^n \lambda_i p_i,$$

a contradiction with $\sum_{i=1}^n \lambda_i p_i \in \operatorname{argmax}_{p \in \operatorname{co}(A)} V(p)$.

2. $\bar{n} > 1$. It follows that $n > 1$. By construction, we have that $p_1 \sim \dots \sim p_{\bar{n}-1}$ as well as $p_{\bar{n}-1} \succ p_{\bar{n}}$. In particular, we have that $\gamma = \sum_{i=\bar{n}}^n \lambda_i \in (0, 1)$. Define $\mu_i = \frac{\lambda_i}{1-\gamma}$ for all $i \in \{1, \dots, \bar{n}-1\}$ and $\gamma_i = \frac{\lambda_i}{\gamma} \in (0, 1)$ for all $i \in \{\bar{n}, \dots, n\}$. We also have that $\sum_{i=1}^{\bar{n}-1} \mu_i = 1$ and $\sum_{i=\bar{n}}^n \gamma_i = 1$. Since \succsim satisfies Betweenness, we can conclude that

$$p_1 \succsim \sum_{i=1}^{\bar{n}-1} \mu_i p_i \succsim p_{\bar{n}-1} \text{ and } p_{\bar{n}} \succsim \sum_{i=\bar{n}}^n \gamma_i p_i \succsim p_n.$$

This implies that $p_1 \sim \sum_{i=1}^{\bar{n}-1} \mu_i p_i \sim p_{\bar{n}-1} \succ p_{\bar{n}} \succsim \sum_{i=\bar{n}}^n \gamma_i p_i$. By Claim 1, we have that

$$p_1 \sim \sum_{i=1}^{\bar{n}-1} \mu_i p_i \succ (1-\gamma) \sum_{i=1}^{\bar{n}-1} \mu_i p_i + \gamma \sum_{i=\bar{n}}^n \gamma_i p_i = \sum_{i=1}^n \lambda_i p_i,$$

proving that $A \ni p_1 \succ \sum_{i=1}^n \lambda_i p_i$, a contradiction with

$$\sum_{i=1}^n \lambda_i p_i \in \operatorname{argmax}_{p \in \operatorname{co}(A)} V(p).$$

Points 1 and 2 prove the implication.

Now we show that (i) implies (ii). For that, let $p \succ q$. Define $A = \{p, q\}$. This implies that $\operatorname{supp}_{\rho}(A) = \{p\}$. Since ρ is a stochastic choice function that admits a Cautious Stochastic Choice representation, it follows that

$$p = \overline{\rho(A)} \in \operatorname{argmax}_{p \in \operatorname{co}(A)} V(p).$$

This implies that $p \succsim \lambda p + (1-\lambda)q$ for all $\lambda \in [0, 1]$. Since p and q were arbitrarily chosen, we thus have showed that for each $\lambda \in (0, 1)$

$$p \succ q \implies p \succsim \lambda p + (1-\lambda)q. \quad (8)$$

By the representation of \succsim , it follows that (8) also holds when $p \succsim q$. The representation of \succsim also implies that it satisfies Convexity, that is, for each $\lambda \in (0, 1)$

$$p \succsim q \implies \lambda p + (1-\lambda)q \succsim q.$$

We conclude that \succsim satisfies Betweenness. ■

Proposition 9. *Let \succsim be a binary relation that admits a Cautious Expected Utility representation \mathcal{W} which is compact and such that each element of \mathcal{W} is concave. If \succsim satisfies Betweenness, then there exists a stochastic choice function ρ that satisfies Regularity and admits a Cautious Stochastic Choice representation \mathcal{W} .*

Proof of Proposition 9. Define $V : \Delta \rightarrow \mathbb{R}$ by

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

Define also ρ to be such that $\rho(A)(q) = \frac{1}{|\text{argmax}_{p \in A} V(p)|}$ for all $q \in \text{argmax}_{p \in A} V(p)$ and for all $A \in \mathcal{A}$. Since \succsim satisfies Betweenness, we have that $\text{co}(A) \ni \overline{\rho(A)} \succsim q$ for all $q \in \text{co}(A)$. In other words, we can conclude that

$$\overline{\rho(A)} \in \text{argmax}_{p \in \text{co}(A)} V(p)$$

yielding that ρ admits a Cautious Stochastic Choice representation \mathcal{W} . At the same time, Regularity is satisfied by construction. \blacksquare

To sum up, Regularity restricts \succsim to be a preference that further satisfies Betweenness, but, at the same time, the previous result shows that it is compatible with any Betweenness preference.

Appendix C: Proof of the Results in the 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. 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 $\rho(A_1) = p$. Consequently, we have pRq . Moreover, if we have $k \in \mathbb{N}$ and A_1, \dots, A_k such that $\rho(A_1) = q$ and $\rho(A_i) \in \text{co}(A_{i-1})$ for $i = 2, \dots, k$, Rational Mixing implies that $p \notin \text{co}(A_k)$. This shows that we cannot have $q \text{tran}(R)p$. We conclude that $\text{tran}(R)$ is an extension of the first order stochastic dominance relation. Now pick any complete extension \succsim of $\text{tran}(R)$. 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 $q \in \text{co}(A)$, for every $A \in \mathcal{A}$. Consequently, we have $\overline{\rho(A)} \succsim q$ for every $q \in \text{co}(A)$, for every $A \in \mathcal{A}$. This proves Theorem 1. \blacksquare

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 \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 $\{p\} = \text{supp}_\rho(\{p, \delta_y\})$. \parallel

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 $V : \Delta \rightarrow [w, b]$ by $V(p) = \sup D_p$ for every $p \in \Delta$.²⁶ 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 $q \in \text{co}(A)$.

Proof of Claim. Fix a choice problem A and $q \in \text{co}(A)$. 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 $V(p) \geq V(q)$. \parallel

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, $\delta_{x+\varepsilon} \in \text{supp}_\rho(\{p^{m_k}, \delta_{x+\varepsilon}\})$.

²⁶We 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.

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 $V(p^m) \rightarrow V(p)$. \parallel

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 Weak Stochastic Certainty Effect 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 $\lambda p + (1 - \lambda)q \succ \gamma(\lambda p + (1 - \lambda)q) + (1 - \gamma)(\lambda \delta_x + (1 - \lambda)q) \succsim \gamma(\lambda p + (1 - \lambda)q) + (1 - \gamma)\delta_z$ for every $\gamma \in [0, 1]$.²⁷ 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 Effect 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 $\lambda p + (1 - \lambda)q \succsim \gamma(\lambda p + (1 - \lambda)q) + (1 - \gamma)(\lambda \delta_y + (1 - \lambda)q)$ 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 $\{\lambda \delta_y + (1 - \lambda)q\} = \text{supp}_\rho(\{\lambda \delta_y + (1 - \lambda)q, \delta_{z'}\})$. By Weak Stochastic Certainty Effect, this implies that $\{\gamma(\lambda p + (1 - \lambda)q) + (1 - \gamma)(\lambda \delta_y + (1 - \lambda)q)\} = \text{supp}_\rho(\{\gamma(\lambda p + (1 - \lambda)q) + (1 - \gamma)(\lambda \delta_y + (1 - \lambda)q), \gamma(\lambda p + (1 - \lambda)q) + (1 - \gamma)\delta_{z'}\})$ 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 $\{\lambda p + (1 - \lambda)q\} = \text{supp}_\rho(\{\lambda p + (1 - \lambda)q, \delta_z\})$. Since z was arbitrarily chosen, we conclude that $V(\lambda p + (1 - \lambda)q) \geq V(\lambda \delta_y + (1 - \lambda)q)$. Since y was arbitrarily chosen and V is continuous, this now implies that $V(\lambda p + (1 - \lambda)q) \geq V(\lambda \delta_x + (1 - \lambda)q)$. \parallel

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

²⁷Recall that we are assuming that $x = w$, for now.

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 $V(p) \geq V(q)$. \parallel

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 $V(\lambda\delta_x + (1-\lambda)\delta_w) = \hat{x} > \hat{y} = V(\lambda\delta_y + (1-\lambda)\delta_w)$. \parallel

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 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 $p, q \in \Delta$,

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

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 $V : \Delta \rightarrow \mathbb{R}$ 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 $\hat{\succsim}$ 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. \blacksquare

Proof of Proposition 1. 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 $q \in \Delta$,

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

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

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

and that $\{\delta_y\} = \text{supp}_\rho(\{q, \delta_y\})$ 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. ■

Proof of Proposition 2. By Proposition 1, 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

$$V(p) = \min_{v \in \mathcal{W}} v^{-1}(\mathbb{E}_p(v)) \quad \forall p \in \Delta. \quad (9)$$

By the proof of Theorem 3, we also know that V in (9) is also represented by the normalized, compact and convex set $\mathcal{W}_{\text{max-nor}}$. By Theorem 2 (and its proof) in (Cerreia-Vioglio et al., 2015a) and since \mathcal{W} is convex and compact, $\mathcal{W}_{\text{max-nor}} \subseteq \text{cl}(\text{co}(\mathcal{W})) = \text{cl}(\mathcal{W}) = \mathcal{W}$. By setting $\widehat{\mathcal{W}} = \mathcal{W}_{\text{max-nor}}$ we obtain the desired conclusion. ■

Proof of Proposition 4. 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 satisfies strict first order stochastic dominance.

We start by proving the following claim.

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

Proof of Claim. 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 $q \succsim p$. By the representation, this means that there exists $\alpha \in (0, 1)$ such that $\alpha(\lambda p + (1 - \lambda)\delta_w) + (1 - \alpha)q \succsim q$, hence $\alpha p + (1 - \alpha)q \succ p, q$. Thus \succsim 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 function. ||

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 $q \in \text{co}(A \cup \{\delta_y\})$ and there exists $p \in \text{co}(A)$ 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. But Cerreia-Vioglio et al. (2015b) shows that if a preference relation that satisfies Betweenness admits a Cautious Expected Utility representation, then this representation must have an infinite set of utilities. Since $|\mathcal{W}| < \infty$, then \succsim must admit a point of strict convexity. By the previous claim, ρ must be a non-degenerate stochastic choice function. \blacksquare

Proof of Proposition 5. Suppose that ρ admits a Cautious Stochastic Choice representation \mathcal{W} and satisfies Regularity. Define $V : \Delta \rightarrow \mathbb{R}$ as

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

Fix a choice problem A and pick a lottery $q \in A$ such that

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

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 Regularity, this implies that $\{q, \delta_x\} \cap \text{supp}_\rho(A \cup \{\delta_x\}) = \emptyset$. But then, if $q \in \text{supp}_\rho(A)$, Regularity would be violated for some alternative $p \in \text{supp}_\rho(A \cup \{\delta_x\})$. We conclude that $q \notin \text{supp}_\rho(A)$. This shows that statement 1 in the proposition is

true. Since Proposition 8 shows that both statements are in fact equivalent, this also gives us statement 2. ■

Proof of Proposition 6. Define $V : \Delta \rightarrow \mathbb{R}$ 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 $\{\lambda p + (1 - \lambda)r\} = \text{supp}_\rho(\{\lambda p + (1 - \lambda)r, \lambda q + (1 - \lambda)r\})$. By applying Proposition 5 again, it follows 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 $u : [w, b] \rightarrow \mathbb{R}$. By 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 (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,” Levine’s Working Paper Archive 81457700000000447, David K. Levine.
- 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 (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. IJIMA, 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. (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.
- 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.
- 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.